## Uvod

SQL(structured query language) je jezik koji se koristi za slanje upita ka RDBMS(relational database management system) koji skladisti podatke po relacionom modelu. Relacioni model je semanticki model za prezentovanje podataka koji se bazira iz dve grane matematike – set theory(teorija skupova) i predicate logic.

T-SQL je dijalekt SQL-a.

### SQL

SQL je ANSI i ISO standard language za queryovanje I storovanje podataka RDBMS-a.

SQL se sastoji iz nekoliko kategorija izraza:  
- **DML** (data manipulation language) (select, update, insert… tipovi izraza za upit i manipulaciju podataka)  
- **DDL** (data definition language) (create, alter, drop tipovi izraza, za manip. definicija podataka)  
- **DCL** (data control language) (grant, revoke… tipovi izraza za manipulaciju dozvola)

Ova knjiga se fokusira na DML tipove izraza, znaci SELECT, INSERT, UPDATE, DELETE, TRUNCATE, and MERGE

Common misunderstanding is that TRUNCATE is a DDL statement, but in fact it is a DML statement

### Set Theory

*By a set we mean any collection M into a whole of definite, distinct objects “m” (which are called the “elements” of M) of our perception or of our thought.*

* *Georg Cantor*

//Mozes malo vise prouciti o skupovima kasnije ako te interesuje, ali osnovna skola

### Predicate logic

Koristi se za upravljanje podacima i izvrsavanje upita nad podacima. PREDICATE je u sustini izraz/svojstvo koje je tacno ili nije tacno(holds, or doesn’t hold)

Relacioni model se oslanja na predicate za zadrzanje logickog integriteta podataka i definisanje strukture istih. U sustini primer bi bio CONSTRAINT koji ce nad tabelom “employees” da postavi pravilo da plata mora da bude > 0 => salary > 0

### The relational model

Relacioni model je sematicki model za upravljanje podacima i manipulaciju koji se bazira na teoriji skupova i logici. Cilj ovog modela jeste omogucavanje konzistentne prezentacije podataka uz minimalnu ili bez imalo redundantnosti gde ne moras da se oslanjas na intuiciju jer jelte, bazira se na matematici.

Relacioni model ukljucuje koncepte poput “propositions, predicates, relations, tuples, attributes & more”.

### Propositions, predicates & relations

Da deo “relacioni” u izrazu “relacioni model” stoji tu na osnovu koncepta “relacija” je netacan. Zapravo potice iz matematickog izraza relacija. U teoriji skupova, relacija je representation(predstava) skupa. U relacionom modelu, relacija je skup (related) informacija. U SQL-u taj skup se naziva tabela(not exact counterpart, but close).

Key point -> relacija jeste jedna tabela.

Operacije nad relacijama na osnovu relacione algebre rezultuju relacijama

//Relational model razlikue relaciju od relacione promenljive but to keep things simple, necemo o tome sada. Koristicemo izraz “relacija” za oba.

Relacija se sastoji iz headinga i bodyja.

Heading jeste spisak atributa(kolona u SQL-u) gde svaka kolona ima svoj naziv i tip podatka koje ce da sadrzi.

Body jeste skup tuple-a(row u SQL-u) gde se svaki element identifikuje kroz key. To KIS, reci cemo da je tabela skup redova.

A diagram of a workflow

AI-generated content may be incorrect.

Keep in mind da elementi u skupu(tuples) i kolone/atributi mozda mogu izgledati kao da imaju redosled. Oni nemaju redosled(order) i ne mogu imati redosled, po definiciji skupova.

To isto vazi i za recorde unutar tabele.

Kada formalizujes neku pojavu u prirodu tako da je mozes prezentovati kroz skup – ti ces u sustini koristiti razne predicates. Za opisivanje kolone/atributa, npr ti ces koristiti TIP(type).

**TYPE** je najjednostavniji oblik predikata. Npr int ogranicava kolonu da vrednost bude u opsegu –2,147,483,648 do 2,147,483,647. Tip nije samo integer/date. To su jednostavni tipovi.

Tip moze biti i kompleksan gde bi kompleksan bio npr neka enumeracija vrednosti.

### Missing values

Jelte, postoji debata o tome da li bi predikat trebao podrzavati dvo-vrednostne ili trovrednosne vrednosti. Dvovrednosni predicate bi bio npr true/false.

Trovrednosni bi bio true/false/null tj nepostojanje vrednosti. Dvovrednosni sistem podrazumeva “the law of excluded middle” sto je matematicko pravilo koje kaze da “nema sredine”. Ili jesi, ili nisi. Sa Trovrednosnim sistemom ti kazes “unknown/ne znam” da je takodje vrednost.

Neki tvrde da se NULL/unknown i trovrednosni predicate logic na ubrajaju u relacioni model.

SQL supports three value predicate logic.  
Postoji “Codd” koji tvrdi da treba da postoji cetvorovrednosni predicate sistem u kom kada postoji nepoznata vrednost(null), ona se deli na “primenjivu” i “neprimenjivu”. Neprimenjiva bi bila kada employee nema mobilni telefon pa onda postojanje atributa mobilnog telefona nema primenu tj non-applicable.

Ukoliko employee ima mobilni telefon onda je “primenjiva” nepoznata vrednost jer jelte, ima telefon samo mi ne znamo koji je. SQL ne podrzava cetvorovrednosni predicate sistem.

### Constraints

Jedna od prednosti relacionog modela jeste mogucnost za definisanjem integriteta podataka kao deo modela. Constraints su pravila koja se jelte definisu u modelu, a koje sprovodi RDBMS. Jedan od najcescih jeste da li atribut dozvoljava ili ne dozvoljava null marker.

Constraints takodje se namecu kroz sam model – npr ako imas tabelu sa 3 kolone, onda neces imati 4 atributa :)

Za sprovodjenje referential integrity koriste se candidate keys i foreign keys. Na osnovu candidate key mozes uniquely identifikovati record u tabeli npr. Tipicno je da se jedan od candidate keys koristi kao primary key.   
U slucaju ljudi to ce biti JMBG npr, a mozes i “id”. Svi drugi candidate keys pored tvog primary key-a se nazivaju alternate keys.

Foreign keys are used to enforce referential integrity. Definise se nad jedan ili vise atributa u relaciji i referencira candidate key u toj relaciji ili nekoj drugoj relaciji.

### Normalization

Relacioni model poseduje pravila normalizacije, kao i u matematici. Naime, sluzi da se smanji redundantnost podataka.

Postoje x3 nominalne forme - 1NF, 2NF, 3NF

1st nominal form(1NF) kaze da svaki record(tuple) u tabeli(relaciji) mora biti unique i da svaka kolona(atribut) mora biti atomican.

2nd nominal form(2NF) ima dva pravila. Prvo kaze da podaci moraju da zadrze njihov prvu normalnu formu. Drugo je vise vezano za kljuceve(candidate & foreign).

Svaki atribut u relaciji sa candidate key, teba da bude funkcionalno zavisan od tog candidate key-a. To znaci da atribut ne moze da bude parcijalno zavisan od candidate key-a.

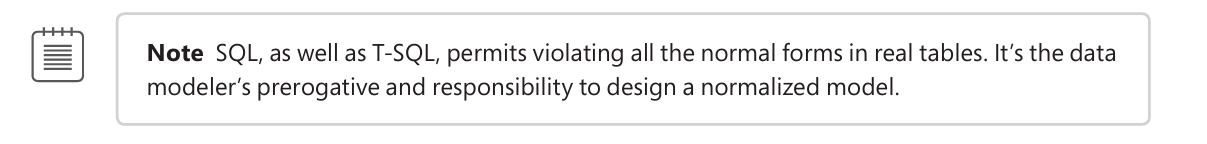
To put it more informally, if you need to obtain any nonkey attribute value, you need to provide the values of all attributes of a candidate key from the same tuple

A screenshot of a computer

AI-generated content may be incorrect.

TLDR ,ne mozes da imas dva primary key-a gde ces na osnovu jednog ili na osnovu drugog raditi pretrage.

3rd nominal form(3NF) takodje ima dva pravila. Prvo je da mora se ispostovati 2nd nominal form. Drugo je da svi non-key atributi moraju biti individualni i nezavisni jedni od drugih(must be dependent on candidate keys nontransitively.)



## Types of database workloads

Dva najcesca workloada su za OLTP(Online transaction processing) i DW(data warehousing).

Za DW to mozes sa AzureSQL/SqlServer koji koristi SMP(symmetric multiprocessing) architecture ILI za malo zahtevnije workloade Synapse Azure Analytics koji koristi MPP(Massive parallel processing)

A diagram of a diagram of a database

AI-generated content may be incorrect.

Here’s a quick description of what each acronym represents:

■ OLTP: online transactional processing   
■ DSA: data-staging area   
■ DW: data warehouse   
■ ETL / ELT: extract, transform, and load; or extract, load, and transform

### OLTP – Online Transaction Processing

Relacioni model jeste primarno “optimizovan” za OLTP workloads koji predstavljaju neki bazicni CRUD kada ti imas online transakcije koje na taj nacin rade sa podacima, znaci primarni cilj ti je data entry bez da imas neki reporting.

Ovde su tabele cesto i normalizovane, ali upravo zbog te normalizacije ovaj use case ne odgovara reportingu jer u OLTP mi cemo imati puno entiteta sa kompleksnim relationshipima.

Ako bi to pokusali na reporting scenariu, mi cemo imati mnogo kompleksnih queryja sa losim performansom.

### DW – Data Warehousing

Ovo je okruzenje koje je optimizovano za pribavljanje podataka i reporting purposes. Ukoliko opsluzuje celu organizaciju onda se to naziva data warehouse. Ako opsluzuje samo departman unutar organizacije ili jedan mali specifican context onda se naziva data mart.

U ovim slucajevima model podataka je optimizovan primarno za retrieval needs.

The model has intentional redundancy, fewer tables, and simpler relationships, ultimately resulting in simpler and more efficient queries than an OLTP environment.

Najjednostavniji data warehouse dizajn se naziva **star schema**. Ona ukljucuje nekoliko dimenzija schema i fact table(tabelu cinjenica). Svaka dimenzija predstavlja jednu temu na osnovu koje zelis da radis reportove – na primer sistem koji radi sa “orders & sales” verovatno ce zahtevati sledece dimenzije: customers, products, employees & time.

Svaka dimenzija je implementirana kao jedna tabela sa redundantnim podacima.

single ProductDim(product dimension) table instead of three normal ized tables: Products, ProductSubCategories, and ProductCategories

Ukoliko uzmes i normalizujes tabelu dimenzije(u vise tabela jelte), onda dobijas nesto sto se naziva **snowflake dimension**. Takva schema koja podrzava normalizovane dimenzije se naziva **snowflake schema**(znaci imas jedan ili vise snowflake, ne mora sve u snowflake).

**Fact table** tj tabela cinjenica sadrzi “facts and measures”(cinjenice i mere) poput kolicina, vrednost za svaku relevantnu kombinaciju kljuceva dimenzija.

Npr za svaku relevantnu kombinaciju customer, product, employee & day – tabela cinjenica ce imati red koji sadrzi quantity & value.

Note – podaci u dimenzijama su cesto “pre agregirani” za razliku od OLTP podataka koji su skladisteni cesto “po transakciji”.

### ETL/ELT

Proces koji povlaci podatke sa izvornih sistema(OLTP & drugi), manipulise im i ucitava ih u datat warehouse se naziva **Extract, transform & load ETL.**

Neki sistemi prvo povlace podatke sa izvornih sistema, ucitavaju ih u data warehouse i na kraju ih transformisu se nazivaju **Extract Load Transform – ELT**

**Microsoft provides an on-premises tool called Microsoft SQL Server Inte gration Services (SSIS) to handle ETL/ELT needs which comes with SQL Server license.**

Microsoft also provides a serverless cloud service for ETL/ELT solutions called **Azure Data Factory**

Cesto procesi koji vrse OLTP -> DW i koji sadrze neki ETL/ELT posao, imaju i DSA**(Data-Staging-Area)** koja moze biti u nekoj npr relacionoj bazi i sluzi kao data cleansing area. Zona gde ce se podaci “preciscavati”.

## SQL Server Architecture

### On-Prem & Cloud flavors of Sql Server.

Inicijalno je tu bio samo jedan RDBMS – Microsoft Sql Server koji je sluzio za on premises baze. Danas unutar svoje “data platforme”, MSFT nudi onprem, box solutions i service based cloud solutions. Mnogo vise :)

### On Premises

Ova nijansa RDBMS-a se naziva SQL Server koji je najtradicionalnija verzija sto msft nudi. Ovde customer vodi racuna o svemu – hardver, patchovi, disaster recovery, data retention i slicno.

Customer moze da instalira i vise instanci SQL servera na istom backendu i radi sta hoce u sustini. Max kontrola.

### Cloud

Cloud computing varijanta nudi cloud compute & cloud storage opcije “po zahtevu” sa deljenog “pool of resources”. MSFT RDBMS mogu biti ponudjene i sa public cloud i private cloud varijanti.

Za private cloud infrastrukturu se misli na cloud infrastrukturu koju jedna organizacija koristi putem tehnologije virtuelizacije. Cesto je hostovana lokalno.

Za public cloud infrastrukturu – ovde microsoft sve resava za tebe tako sto svi servisi public cloud infrastrukture su ti dostupni preko “network” tj mreze.

Ovde imamo x2 opcije, a to su IaaS i PaaS.

**IaaS** je kada hostujes SQL Server na VM. Ovde je na tebi da podesavas sve osim underlying hardware, jel tako. It’s essentially like maintaining your own SQL Server installation—one that happens to reside on Microsoft’s hardware.

**PaaS** je ukoliko koristimo microsoft managed sql server instance. Hardware, software installation and maintenance, high availability and disaster recovery, and patching are all responsibilities of Microsoft.   
Customer jedino o cemu treba da brine je index & query tuning.

### SQL Server Instances

U sustini ovo su instalacije SQL Servera na jednom on prem servisu. Svaka instalacija(proces) ne deli nista drugo izmedju sebe osim hardvera. Kompletno je nezavisna.

A diagram of a server

AI-generated content may be incorrect.

Kada imas vise instanci na istom serveru, samo jedna instanca moze biti default instanca. Sve ostale moraju biti named instances. O tome koja instanca je default, odlucuje se na kreiranju same instance i ne moze se kasnije promeniti.

Da bi se client povezao na instancu:  
**(default)** – **{pc\_ip/pc\_name} (Server1)**  
**(named\_instance)** – **{pc\_ip/pc\_name}\{instance\_name} (Server1\Inst2)**

### Databases

Baze(databases) su kao containeri koji sadrze razlicite objekte poput tabela, viewova, stored procedura i drugih objekata. Svaka instanca SQL servera moze sadrzati vise baza.

Setup proces takodje automatski postavi nekoliko sistemskih baza za interne mssql potrebe.

Ti naravno mozes postaviti svoje user databases koje ce sadrzati podatke koje ti zelis.

A diagram of several columns

AI-generated content may be incorrect.

Sistemske baze:

* **master** – Sadrzi “instance wide” metadata informacije, konfiguracije servera, informacije o svim bazama unutar instance i informacije inicijalizacije.
* **model** – Ova baza se koristi kao template za nove baze. Svaka nova baza koju kreiras incijalno je kreirana kao kopija model baze, pa ako zelis neke objekte poput user defined data types da budu dostupni u svim novokreiranim bazama ili da novokreirane baze budu unapred konfigurisane na neki nacin, to ovde mozes da podesis. Note, dodatne izmene u “model” bazi se nece odraziti na vec postojece baze
* **tempdb** – ovde SQL Server skladisti privremene podatke poput work tables, sort & hash table data kada treba da se persistuju, row versioning information i slicno. SQL Server ti dozvoljava da ovde kreiras i svoje tabele, ali **tempdb** se unistava i re-kreira kao kopija **model** baze svaki put kada restartujes instancu SQL servera.
* **msdb** – Ovu bazu koristi najcesce SQL Server Agent servis da skladisti svoje podatke. Njegov zadatak je automacija u koje spadaju “jobs, schedules, and alerts”. SQL Server Agent takodje vrsi replikacije. Pored toga, msdb sadrzi podatke o SQL Server features poput database mail, service broker, backups, and more.
* **resource** – resource baza je skrivena, read only db koja cuva definicije o svim sistemskim objektima. Kada queryujes system objekte u bazi, ispada kao da su oni u **sys schema** lokalne baze, ali u stvarnosti sve te definicije se nalaze u resource bazi.

Mozes kreirati 32.767 user baza unutar instance. User database skladisti objekte i podatke aplikacije.

Postavljanjem “collation” propertya na db level ce ti odluciti o default language support, case sensitivity & sort order za character data u bazi. Ako ne specificiras, uzimaju se default collation instance.

* Security concepts

Naime kako bi pristupio instanci moras imati pristup ti kao user. To je login.   
Potom kada zelis da pristupis nekoj bazi, db admin mora da namapira tvoj login na database usera.

Database user je entitet na kojeg se nadodaju permisije ka db objektima. SQL Server supportuje **contained databases** feature koji “breaks the connection” izmedju db usera i instance level logina. User(Windows/SQL Authenticated) je fully contained unutar odredjene baze i nije povezan u login na instance level.

To znaci da kada se povezujes na SQL Server kao user, moras da specificira bazu na koju se povezujes. Ne mozes kasnije switchovati baze.

Kada radis sa Azure SQL Managed instance, ti se brines jedino o logickom sloju. Ako radis sa SQL Serverom na VM ili on-prem, onda moras da radis i sa fizickim slojem baze. To znaci da radis i sa sledecim:

A diagram of a computer system

AI-generated content may be incorrect.

Baza se sastoji iz data files, transaction log files, opciono checkpoint files koji sadrze memory optimized data(In-Memory OLTP feature).

Svaka baza mora da ima barem jedan data file i barem jedan log file. Data files cuvaju object data, a log files informacije koje SQL Serveru trebaju da bi odrzao transakcije.

Pri kreiranju baze mozes definisati razlicita svojstva za data i log files ukljucujuci i file name, location, initial size, max size i autogrowth increment.

Iako SQL Server moze da upisuje u vise data files u paraleli, sa log fileovima moze samo jedan(one at a time) u sekvencijalnom pristupu.

**Datafiles** su organizovani u **filegroups** koje su u sustini logicke grupacije. Filegroup je target za kreiranje objekta poput table/index. Informacije o objektu ce biti rasporedjene kroz fileove koji pripadaju target file groupi.

Filegroups su tvoj nacin za kontrolisanje fizickih lokacija tvojih objekata. Database mora posedovati barem jednu filegroup koja se naziva **PRIMARY** i opciono moze imati druge filegroups.

Primarni data file(**.mdf** ekstenzija) se nalazi u primary filegroup i to je pirmarni data file za bazu i system catalog. Opciono mozes dodati secondary data files(**.ndf** ekstenzija) unutar PRIMARY.

User filegroups sadrze samo secondary data files. Mozes odabrati koja filegroup je marked kao default filegroup.

Kada object creation statement ne govori preciznije o tome gde se kreira, on ce se kreirati u default filegroup.

.mdf -> master data file

.ldf -> log data file

.ndf -> not master data file

SQL Server db engine u sebi ima takodje i memory-optimized engine “In-Memory OLT”. Uz taj feature mozes integrisati memory optimized objects poput memory optimized tables i natively compiled modules(procedures, functions & triggers) u bazu. Da bi to uradio moras da kreiras filegroup u bazi koji ces markovati kao containing memory optimized data i unutar njega bar jedan path u folder. Sql server storeuje checkpoint files u memory optimized data u tom folderu i koristi ih da recoveruje data svaki put kad je sql server restartovan.

### Schemas & Objects

Da je baza container objekata je malo simplified verzija. Baza sadrzi schemas. Schemas sadrze objekte. Schema je container(namespace) objekata poput tabela, views, stored procedures & drugo.

A diagram of a database

AI-generated content may be incorrect.

Permisije mozes kontrolisati na schema nivou. Npr useru mozes dozvoliti SELECT permisija za samo jednu schemu.

Pri svojim upitima ukoliko ne ukljucis schema name, SQL Server ce apply process za resolve schema name poput proveravanja da li objekt postoji u **user default schema**, a ako ne postoji onda proverava da li postoji u **dbo** schema.

Ukoliko isti objekat postoji u vise schema onda je bolje da specificiras schema pre nego pokusas da ga referujes jer mozda dobijes razlicite rezultate

## Creating tables and defining data integrity

Ovi primeri ovde se odnose na **dbo** schema koja se automatski kreira u svakoj bazi i postavlja za default schema za usere osim ukoliko nije drugacije naznaceno.

USE TSQLV6; --> sets current db context

DROP TABLE IF EXISTS dbo.Employees; --> mozes i da ni ne kreiras ako postoji vec

CREATE TABLE dbo.Employees --> **dbo.Employees** two part name(msft recommended)  
(   
empid INT NOT NULL,   
firstname VARCHAR(30) NOT NULL,   
lastname VARCHAR(30) NOT NULL,   
hiredate DATE NOT NULL,   
mgrid INT NULL,   
ssn VARCHAR(20) NOT NULL,   
salary MONEY NOT NULL   
);

Ako ommitujes schema name onda ide u default schema.  
Pri create table specificiras <attribute\_name> <data\_type> <nullability>

### Defining data integrity

Imamo procedural data integrity gde se integritet podataka cuva kroz stored procedure i trigere i declarative data integrity gde se integritet cuva kroz samu definiciju modela.

declarative data integrity: primary key, unique, foreign key, check, default constraints, nullability, data type.

Takve constraintove mozes u toku CREATE TABLE definisati, a i dodati kasnije kroz ALTER TABLE. Svi constraintovi osim default constrainta se mogu definisati kao composite constraints.

### Primary key constraints

PK Constraint forsira uniqueness of rows i nedozvoljava NULL u constraint atributima. Svaka kombinacija vrednosti atributa se znaci moze pojaviti samo jednom(unique).

Ako pokusas da dozvolis null pod PK, dobijes error.

ALTER TABLE dbo.Employees  
 ADD CONSTRAINT PK\_Employees  
 PRIMARY KEY(employeeid)

Behind the scenes, sql server kreira index da bi naterao uniqueness logickog primary key constrainta. Unique index je fizicki objekat kojeg sql server koristi da enforceuje uniqueness.

Indexi takodje se koriste da bi se speed up queries da bi se izbegao sorting ili nepotrebni full table scans.

### Unique constraints

Enforceuje uniqueness row-a. Dozvoljava ti implementaciju koncepta alternate key-a.

Mozes imati vise unique constraintova na atributima tabele. Kolona moze biti i null, nije restricted samo na not null kolone.

ALTER TABLE dbo.Employees  
 ADD CONSTRAINT UNQ\_Employees\_ssn  
 UNIQUE(ssn)

Behind the scenes kreira se isto unique index.   
Jedna interesantna stvar, po SQL-u unique se definise kao “Unique constraint na T je zadovoljena akko ne postoje dva reda R1 i R2 od T takvih da R1 i R2 imaju istu ne null vrednost”.

Znaciu unique se ubraja samo za vrednosti. Kako je null nepostojanje vrednosti, mi trebali imati vise redova sa unique attributom gde je null “vrednost”.

To je kako SQL standard definise null handling unique constraintova.

Po SQL Serveru ti ako stavis unique na nullable atribut mozes imati samo jedan null od svih redova.

Ukoliko zelis unique po SQL standardu:

CREATE UNIQUE INDEX idx\_ssn\_notnull ON dbo.Employees(ssn) WHERE ssn IS NOT NULL

### Foreign key constraints

Foreign key razresafa referentni integritet. Poenta FK je da ogranici vrednosti u FK koloni na one koje postoje u referenced columns.  
  
CREATE TABLE dbo.Odrders  
(  
orderid INT NOT NULL,  
custid INT NOT NULL,  
desc VARCHAR(20),  
qty INT NOT NULL,  
CONSTRAINT PK\_Orders  
 PRIMARY KEY(orderid)  
)

I sad da ubacis FK:  
ALTER TABLE dbo.Orders  
 ADD CONSTRAINT FK\_Orders\_Employees  
 FOREIGN KEY(empid) REFERENCES dbo.Employees(empid)

Postoje “referencing” i “referenced” tabele u FK. Referencing je one koje pozivaju referenced tabelu. Referencing je orders. Referenced je employees.

Nulls su dozvoljeni u FK po defaultu. Cak iako PK ne dozvoljava NULL.

Ovako gore opisani FK je samo za enforceovanje referential action-a koji se naziva “NO ACTION”.

To znaci da ukoliko pokusas da obrises row iz referenced table ili update referenced candidate key attributes, to ce biti rejected ako related rows postoje u referencing tabelama.

Mozes definisati FK sa actions.

Postoje opcije ON DELETE i ON UPDATE sa akcijama poput CASCADE, SET DEFAULT i SET NULL.

CASCADE – operacije ce se cascadeovati u related rows. ON DELETE CASCADE znaci da ce se i delete preneti.

SET DELETE i SET NULL – operacije koje u referenced tabeli menjaju candidate key atribute, postavice vrednost u referencing tabelama na default/null vrednost. Ovde ce postojati potom “orphan rows”

Parent rows with no related child rows su ovek allowed.

### CHECK Constraints

Definises predicate koji row mora da ispostuje da bi bio dodat/modifikovan.

ALTER TABLE dbo.Employees  
 ADD CONSTRAINT CHK\_Employees\_salary  
 CHECK(salary > 0.00);

Modifikacije se priznaje samo kada je evaluacije TRUE ili UNKNOWN. Npr salary od -1000 ce biti rejected, ali 50000 i NULL bice accepted. Ako je null dozvoljen.

### Default constraint

Vezuje se za neki konkretni atribut. Koristi se kako bi osigurao vrednost nekog atributa ukoliko ne bude bio prosledjen pri kreiranjnu reda.

ALTER TABLE dbo.Orders  
ADD CONSTRAINT DFT\_Orders\_orderts  
DEFAULT(SYSDATETIME()) FOR orderts;

Default expression invokeuje SYSDATETIME fn koja uzvraca sa current date i time vrednosti.

## Single Table Queries

### SELECT Statement

Ovaj statement sluzi za vrsenje upita nad tabelom, primenjivanje logickih manipulacija i vracanje rezultata. Ovde ce biti opisan i logicki redosled po kom se razliciti query upiti procesiraju(logical query processing).

DB Engine ima svoj query optimizer koji cesto dodaje dosta transformacija i precica u “fizickom procesiranju queryja” kao deo optimizacije.

SELECT empid, YEAR(orderdate) as orderyear, COUNT(\*) AS numorders   
FROM orders  
WHERE custid = 71  
GROUP BY empid, YEAR(orderdate)  
HAVING COUNT(\*) > 1  
ORDER BY empid, orderyear

U vecini programskih jezika bi se stvari izvrsavale redosledom kako ih ispises, ali u SQL je malo drugacije. U gore queryju je redosled sledeci(logical order):  
1. FROM (queries the rows from the orders table)  
2. WHERE (filters only orders where custid is 71)  
3. GROUP BY (groups the orders by employee id and orderdate)  
4. HAVING ( filters only groups(emp id and ord yr) having more than one order)  
5. SELECT ( returns emp id, ord yr & num of orders for each group)  
6. ORDER BY (sorts the rows in the output)  
  
Iako sintaksno ide prvo “select”, on se skoro poslednji izvrsava

### FROM clause

U ovom clause specificiras imena tabela nad kojima zelis da vrsis upit.

**FROM Sales.Orders**

Kada selektujes sve recorde iz tabele, ne moras da delimitujes identifiere(kolone).  
Ovako nesto: **SELECT id, name, surname FROM users** je sasvim okej.

Medjutim ukoliko je tvoja kolona “reserved keyword”, pocinje s brojem, sadrzi razmak i slicno – onda koristis T-SQL specific formu za delimiting ili navodnike:  
**SELECT id, “Order Details”, [Full Name] from [dbo].[Users]**

### Where Clause

Definises predikat ili logicki izraz na osnovu kog se filtriraju rezultati iz FROM clause-a. Oni redovi koji su evaluirani sa “true” su vraceni, a ostali odbaceni. Pod ostali to su “FALSE” i “UNKNOWN”. Zapamti – three-valued predicate logic.

WHERE izraz moze mnogo da utice na performans, na primer ako u query-u uzmes izraz koji koristi indexe, onda ces mnogo brze filtrirati rezultate nego da se radi full table scan.

### GROUP BY Clause

Mozes koristiti GROUP BY kako bi logicki rasporedio redove u grupe. Grupe su odredjene na osnovu elemenata/izraza u samom GROUP BY-u.

Npr: **GROUP BY empid, YEAR(orderdate)**

Ovaj gore izraz znaci da ce se u group by phase kreirati grupa za po svaku distinct vrednost empid i godine.

Ukoliko koristis GROUP BY, onda se izrazi SELECT, HAVING i ORDER BY odnose na grupe, a ne individualne redove.

Elements that do not participate in the GROUP BY clause are allowed only as inputs to an aggregate function such as COUNT, SUM, AVG, MIN, or MAX

A screenshot of a computer screen

AI-generated content may be incorrect.

Count npr vraca broj redova u svakoj grupi.

Sum npr vraca sumu svih freight values unutar grupe.

NOTE: Sve aggregate funkcije ignorisu “NULL” tj nepostojanje vrednosti

Na primer da imamo grupu sa redovima u kojima je jedna kolona null(npr da imamo 5 redova u grupi), onda ce COUNT(\*) vratiti 5.   
Ukoliko radimo npr COUNT(by\_col\_with\_null) onda ce vratiti 4.   
  
COUNT(\*) broji samo redove dok ce COUNT(qty) brojati kolone sa vrednostima.

Prebrojavanje unikatnih vrednosti po COUNT-u:  
COUNT(DISTINCT qty) ili npr SUM(DISTINCT qty)

* Funkcije MAX & MIN traze po grupi

Ukoliko zelis da trazis po redovima najvecu ili najmanju vrednost u koloni onda mozes da koristis funkcije LEAST & GREATEST

### HAVING Clause

Ovo je group filter. Jedino grupe u kojima se HAVING equateuje sa “true” ce biti prikazane/vracene. False/unknown je discarded.

HAVING clause se procesira tek nakon sto su redovi grupisani pa mozes i koristiti aggregate fn’s u filteru.   
Npr grupe sa samo preko 5 porucenih artikala:  
HAVING COUNT(qty) > 5

ili grupe koje imaju vise od jednog reda u sebi

HAVING COUNT(\*) > 1

### SELECT Clause

Specificiras atribute/kolone koje zelis da se vrate u result tabeli query-a.

Kad pises SELECT \* FROM <tabela>, taj \* ce selektovati kolone redosledom koje su napisane u CREATE TABLE izrazu. To je cesto losa praksa i uvek treba da specificiras sta zelis da selektujes tj kolone. Ako izmenis taj redosled nekako, onda ce u transakcionom programu taj \* znaciti drugu stvar i mozes dobiti greske.

A screenshot of a document

AI-generated content may be incorrect.

### ORDER BY Clause

Sluzi da sortira redove rezultata upita. Koristi se za output purposes.

Logicki, izvrsava se posle SELECT statementa. Znaci SELECT je predposlednji, a ORDERBY poslednji.

Posto se redosled clanove ne moze garantovati u tabelama kad se vraca rezultat, a ti zelis garanciju onda moras koristiti ORDER BY. Medjutim, ordered redovi se nazivaju “cursori”. Sad kako je redosled zagarantovan, tako je i to drugi tip rezultata u pitanju koji ne moze da se “qualify” kao tabela.

U ORDER BY mozes da se pozivas na aliase iz SELECT statementa. To cak ni u samom SELECT statementu ne mozes. Ni u jednom drugom statementu jer jelte, redosled operacija. Ovo ne moze:

SELECT name as xname, UNIQUE xname  
 FROM tabelica   
WHERE xname LIKE ‘%kita%’

ASC – ascending order by(default)  
DESC – descending order by

U ORDER BY mozes specificirate i kolone koje nisu specificirane u SELECT statementu  
SELECT name   
FROM users  
ORDER BY age

However, when the DISTINCT clause is specified, you are restricted in the ORDER BY list only to elements that appear in the SELECT list.

### The TOP and OFFSET-FETCH filters

Ovo su x2 filtera koji se ne baziraju na predicateu poput WHERE/HAVING nego se baziraju na osnovu broja redova i redosleda(rows & ordering)

**TOP filter**

Sluzi da ogranicis broj ili procenat redova koje tvoj query vraca. Oslanja se dakle na sam broj/procenat redova koje vraca i redosled(ordering).

**SELECT TOP(5) orderid FROM Sales.Orders ORDERBY orderdate DESC**

TOP Filter je handled posle distinct. To znaci da ako DISTINCT je specificiran u SELECT clause, top filter je evaluated nakon sto su duplicate rows “razreseni”(removed)

Ukoliko koristis “PERCENT” keyword, onda mozes da uzmes npr top 1% redova(will be rounded up)

**SELECT TOP(1) PERCENT orderid FROM Sales.Orders ORDER BY orderdate DESC**

Npr ako imas 830 redova, rezultat ce biti 9 redova jer 1% rounded up nad 830 je 9.

**The OFFSET-FETCH filter**

Problem TOP filtera je sto nije standard i nema “skip” opciju. Enter offset fetch(korisno za paging). OFFSET-FETCH po SQL standardu je considered kao ekstenzija ORDER BY clause-a.

OFFSET clause – koliko redova zelis da skippujes  
FETCH clause – koliko redova zelis da filtriras nakon skipa

SELECT ordereid, orderdate  
FROM Sales.Orders  
ORDER BY orderdate, orderid  
OFFSET 50 ROWS FETCH NEXT 25 ROWS ONLY

Orderuje prvo po orderdate pa orderid, skippuje 50 redova i uzima sledecih 25.  
Uslov za OFFSET-FETCH je da se koristi ORDER BY clause

Takodje FETCH ne mozes da koristis bez OFFSET :)   
OFFSET bez fetcha je allowed, a ako zelis da uzmes samo “X” redova bez skippovanja onda:  
OFFSET 0 ROWS FETCH NEXT 25 ROWS ONLY

Imas i “sinonima”, znaci umesto “NEXT” mozes da kazes “FIRST”  
Umesto “ROWS” mozes da kazes “ROW”. Ista funkcionalnost, samo drugaciji izraz. Cilj je da bude more english like.

### Window functions – quick look

Window function je funkcija koja za svaki red u underlying query, operates on a window(set) of rows koji deriveuju iz underlying query resulta i computuje skalarni(jedan) result value.

Window of rows je definisan sa OVER clauseom. Window funkcije su profound(raznolike).

Bilo bi prerano ulaziti u detalje sada, evo nekih.

ROW\_NUMBER window function.

Znaci, window funkcija radi na skupu redova koji su exposeovani kroz “OVER” clause. Za svaki row, OVER clause exposeuje funkciji podskup redova iz glavnog query resula.

OVER clause moze da restrictuje redove u window-u koriscenjem optional window partition clause(PARTITION BY). Takodje mozes definisati ordering za calculation koriscenjem window order clause(ORDER BY) koji se ne vezuje za query-ev ORDER BY.

Primer:  
SELECT orderid, custid, val,   
 ROW\_NUMBER() OVER(PARTITION BY custid   
 q ORDER BY val) as rownum  
FROM Sales.OrderValues  
ORDER BY custid, valA screenshot of a computer

AI-generated content may be incorrect.

ROW\_NUMBER ce za svaku particiju generisati unique sekvencialno incrementing integere koji predstavljaju redove u rezultatu po particiji. Dakle particija je kao neki interni group by i tu dobijas “podskup”.

Kreira se dakle separate particija za svaki distinct customer id value, pa s toga row number su unique za svakog customera. OVER clause takodje definise ordering u window na osnovu val atributa, pa ce sequential row numbers biti incremented sa particijom na osnovu vrednosti atributa.

NOTE:  
Expressioni u SELECT se evaluiraju PRE distinct clause-a(ako postoji distinct). Iz tog razloga ce se prvo WINDOW pa onda DISTINCT na kraju.

## Predicates & Operators

U sustini “Predicate” je izraz koji vraca TRUE/FALSE/UNKNOWN. Moze se redjati u kombinaciji sa NOT, AND ili OR operatorima.

Ovde prica o IN, BETWEEN i LIKE predicateima.

TSQL supports operatore: =, >, <, >=, <=, <>, !=, !>, !<. Zadnja x3 nisu standardna. != je zapravo <>  
Recommendation je ne koristiti nonstandard opratore.

Od aritmetike: +, -, \*, /, %

Mozes mesati data types i T-SQL ce operand sa vecom “data type precedence” uzeti i staviti kao data type skalarnog izraza.

Dakle x2 integera 5/2 racaju 2, a ne 2.5.

Zato moras uraditi cast:  
CAST(col1 AS NUMERIC(12,2)) / CAST(col2 AS NUMBERIC(12,2))

NUMERIC(12,2) bi bio precision od 12 i 2 mesta posle tacke.

Ako uradis neki INT/NUMERIC tj 5/2.0 onda ces dobiti 2.5 kao rezultat

**OPERATOR PRECEDENCE:**  
A screenshot of a computer

AI-generated content may be incorrect.

## CASE Expressions

CASE expression je scalar expression koji braca vrednost na osnovu “conditional logic”.

CASE nije expression niti statement, nece kontrolisati flow koda. Znaci samo vraca vrednost. Kako je scalar expression onda se moze koristiti gde kog su i skalarni izrazi dozvoljeni: SELECT, WHERE, HAVING, ORDER BY i CHECK constraints

Postoje x2 tipa CASE expressions:  
- **Simple**   
- **Searched**

**Simple** je za poredjenje vrednosti/izraza sa listom mogucih vrednost i vracanje vrednosti na prvom matchu. Ako nema, onda se vraca vrednost u ELSE clause. Ukoliko nema ELSE, onda se defaultuje na ELSE NULL.

SELECT supplierid, COUNT(\*) AS numproducts,  
 CASE COUNT(\*) % 2   
 WHEN 0 THEN ‘Even’   
 WHEN 1 THEN ‘Odd’  
 ELSE ‘Unknown  
 END AS coutparity   
FROM Production.Products GROUP BY supplierid

**Searched** case expression dozvoljava definisanje predikata u WHEN clause umesto da budes ogranicen samo na equality comparisons.

CASE   
 WHEN val <1000.00 THEN ‘Kita’  
 WHEN val >= 1000.00 THEN ‘Velika kita’  
ELSE ‘Uknown  
END AS valuecategory

## NULLS

Dakle predicate logic pored TRUE i FALSE moze da evaluateuje i “UNKNOWN”. Na primer u slucaju NULL-a ce predicate salary > 0 evaluateovani “UNKNOWN” kada je salary null. Ukoliko je iznad 0 onda TRUE, a ako je ispod 0 onda FALSE.

Razliciti statementi razlicito na ovo gledaju jer UNKNOWN nije false.

Filteri poput WHERE i HAVING imaju “accepts TRUE” treatment. To znaci da u WHERE samo oni koji u predicate vracaju TRUE ce biti uzeti u obzir.

CHECK constraint ima statement “rejects FALSE” sto bi znacilo da bi i UNKNOWN i TRUE bio prihvacen.

Obrati paznju na to.

Ukoliko uradis negaciju na unknown, ti i dalje dobijas “unknown”.   
NOT (salary > 0) -> NOT UNKNOWN == UNKNOWN.

NULL = NULL ce takodje evaluateovati na UNKNOWN.

NULL predstavlja odsustvo vrednosti pa iz tog razloga imas operatore:  
IS NULL i IS NOT NULL, which you should use instead of = NULL and <> NULL.

SQL Server 2022 unosi promenu koja se naziva **IS [NOT] DISTINCT FROM** clause koja tretira NULL’s kao values i za rezultat dobijas iskljucivo true/false bez unknown.

<operand1> IS NOT DISTINCT FROM <operand2> - ovo se evaluira kao (=) sto evaluira u TRUE kada se porede dve null vrednosti. Ako je NULL sa non NULL onda je FALSE.

<operand1> IS DISTINCT FROM <operand2> - ovo se evaluira kao (<>) sto evaluira u FALSE kada se porede dve NULL vrednosti. Ako je Non null i Null onda je TRUE.

**SELECT custid, country, region, city   
FROM Sales.Customers   
WHERE region IS NOT DISTINCT FROM N'WA';**

je ideticno sa ==

**SELECT custid, country, region, city   
FROM Sales.Customers   
WHERE region = N'WA';**

Koriscenje two valued modela ima smisla kada poredis dve kolone ili kolone sa konstantama ili npr u stored procedurama/trigerima.

Ranije si morao pisati ovako nesto:  
**SELECT custid, country, region, city   
FROM Sales.Customers   
WHERE region = @region   
 OR (region IS NULL AND @region IS NULL);**

Ukoliko je @region NULL, onda je ovo jedini nacin da dobijes TRUE na NULL = NULL ako je i region column u trenutnom row-u na NULL.   
Da bi to izbegli mozemo napisati:

**SELECT custid, country, region, city  
FROM Sales.Customers  
WHERE region IS NOT DISTINCT FROM @region**

Kada su NULL IS NOT DISTINCT FROM NULL, onda je rezultat TRUE

Isto vazi i za <> i IS DISTINCT FROM; tj razlicistost.

‘Kita’ IS DISTINCT FROM NULL ti vrati “true”

Kako se drugi elementi SQL-a ponasaju prema NULLovima?  
**GROUP BY** – grupise sve NULL-ove zajedno  
**ORDER BY** – sortira sve nullove zajedno; u T-SQL u ascending orderu prvo idu null, onda vrednosti  
**UNIQUE** – po standardnom SQL-u UNIQUE enforceuje unique na non null values. T-SQL gleda na NULL kao vrednost pa mozes samo x1 null imati.

## GREATEST and LEAST functions

Od sql server 2022 mozes stavljati najvece i najmanje vrednosti redova:  
However, sometimes you need to apply maximum and minimum calculations across columns, or across a set of expressions, **PER ROW**

**SELECT orderid, requireddate, shippeddate,   
 GREATEST(requireddate, shippeddate) AS latestdate,   
 LEAST(requireddate, shippeddate) AS earliestdate   
FROM Sales.Orders   
WHERE custid = 8;**

**A number and date of a company

AI-generated content may be incorrect.**

**NULL inputs are ignored, but if all inputs are NULL, the result is NULL.**

Ovo znaci radi poredjenje u samom ROW-u:  
**GREATEST(val1, expression\_2, val3) i onda vraca najvecu od tih.**

Ranije:  
A screenshot of a computer error

AI-generated content may be incorrect.

## All at once operations

Ovo znaci da T-SQL podrzava feature kada svi izrazi koji se pojave u istoj fazi “logickog query processinga” su evaluirani u isto vreme.

Svi izrazi koji se pojavljuju u istoj fazi su tretirani kao “set” i kao ranije pomenuto – set/skup nema redosleda.

This concept explains why, for example, you cannot refer to column aliases assigned in the SELECT clause within the same SELECT clause

Drugi fuckup je short-circuit expressiona sto T-SQL podrzava, ali zbog all at once operations, mozda ti se evaluation ne dogodi kako mislis da treba.

**SELECT col1, col2   
FROM dbo.T1   
WHERE col1 <> 0 AND col2/col1 > 2;**

Dakle, ako je col1<>0 FALSE, onda se ovaj drugi deo nece uzimati u obzir. Ali mozda se prvo evaluira col2/col1 > 2 i dobijes error ako je col1 = 0. Nema divide by zero u matematici :)

SQL Server cesto uzima u obzir estimacije po pitanju kompleksnosti izraza i na osnovu toga moze neki izraz ranije, a neki kasnije da evaluira.

U ovom gore slucaju kako bi smo izbegli failure mozemo koristiti **CASE** expression sa **WHEN** jer je tu redosled zagarantovan.

**SELECT col1, col2   
FROM dbo.T1   
WHERE   
 CASE   
 WHEN col1 = 0 THEN ‘no’  
 WHEN col1 <> 0 THEN ‘yes’  
 ELSE ‘no’  
END ‘yes’**

Zbog kompleksnosti tog workarounda imamo bolju opciju

**SELECT col1, col2 FROM dbo.T1   
WHERE (col1 > 0 AND col2 > 2\*col1) OR (col1 < 0 AND col2 < 2\*col1);**

Think about this :)

## Working with CHARACTER data

Postoje x2 tipa character data:   
- regular  
- n kind

Regular su npr CHAR ili VARCHAR  
n kind su NCHAR i NVARCHAR

N je “national”, to je kao znacenje tog prefixa.

Razlika je da N kind podrzavaju UTF-16 dok regular podrzavaju samo UTF-8.

Kada navodis velicinu karaktera npr VARCHAR(10) ti kazes da tu moze da se smesti 10B karaktera. U teoriji bi to znacilo 10 karaktera, ali u praksi neki karakter moze da zauzima i 2B ili vise prostora.

Kada referujes na regular type literal onda koristis single quotes ‘this is a literal’  
Kada referujes na n kind type literal onda koristis “N” prefix pre quotes N’this is a literal’

Character data types koji nemaju “VAR” u sebi, oni podrzavaju samo onoliko prostora koliko im kazes. Fixed size su. Iz tog razloga su bolji za writefocused systems, ali nisu pogodni toliko za read focused systems jer storage consumption nije toliko optimalno.

NVARCHAR(25) ce mu staviti 25B prostora + 2B za offset data ali u praksi ce user data odluciti o tome koliko prostora ce se tu zauzeti jer je varijabilno jelte. Because storage consumption for these data types is less than that for fixedlength types, read operations are faster.

Ako stavis velicinu na (MAX) to je onda 2GB prostora. Any value with a size up to a certain threshold (8,000 bytes by default) can be stored inline in the row. Any value with a size above the threshold is stored external to the row as a large object (LOB)

### Collation

Collation je svojstvo charater data koje enkapsulira nekoliko aspekata:

* Sort order
* Case sensitivity
* ...jos :)

Da dobijes skup supported collationa i njihovih opisa onda pozivas fn\_helpcollations  
**SELECT name, description   
FROM sys.fn\_helpcollations();**

**A screenshot of a computer

AI-generated content may be incorrect.**

Mozes da podesavas COLLATION izraza(querya) koriscenjem **COLLATE** clause

**SELECT \* FROM Customers  
WHERE lastname COLLATE Latin1\_General\_CS\_AS = N’Davis’**

I ovo ce collate na case sensitive jer “CS”. Ako nemas collate, onda ce se assumeovati default collation.

### Operations and Functions nad characters

**Konkatonacija**: + sign operator, CONCAT, CONCAT\_WS  
For **other operations**: SUB STRING, LEFT, RIGHT, LEN, DATALENGTH, CHARINDEX, PATINDEX, REPLACE, TRANSLATE, REPLICATE, STUFF, UPPER, LOWER, RTRIM, LTRIM, TRIM, FORMAT, COMPRESS, DECOMPRESS, STRING\_SPLIT, STRING\_AGG

### COALESCE FN

Standard SQL kaze da konkatonacija sa “NULL” yielduje sa “NULL” kao rezultatom.   
region + N’ ‘ + city => NULL ukoliko je i jedna od region/city NULL  
Kako bi se NULL tretirao kao empty string onda se koristi **COALESCE function.**

**SELECT custid, country, region, city,   
 country + COALESCE(N',' + region, N'') + N',' + city AS location   
FROM Sales.Customers;**

Coalesce funkcija kad naidje na prvu null, onda samo returnuje empty string.

Potomo imamo CONCAT

### CONCAT FN

Prima se lista charactera i automatski zameni NULL sa empty string   
CONCAT(‘Jure’, N’ je‘, NULL, ‘ derpe’) => Jure je derpe

### CONCAT\_WS

Ovo je konkatonacija sa separatorom. Prvi character je separator. Svaki sledeci character ce se razdvojiti tim separatorom

CONCAT(‘x’, ‘Jure’, N’ je‘, NULL, ‘ derpe’) => Jure x je x derpe; valjda ovako

### SUBSTRING

SUBSTRING(string, start, length)

SELECT SUBSTRING('abcde', 1, 3); => ‘abc’  
Ne ide od 0 start. Takodje ako je length > od lengtha stringa, onda ce se jednostavno vratiti sve bez da se baca neki error

### LEFT & RIGHT

Substring s leva ili desna  
LEFT(‘string’, length) & RIGHT(‘string’, length)

RIGHT(‘string’, 3) => ing;

### LEN & DATALENGTH

**LEN(‘string’)** vraca length stringa

**DATALENGTH(‘string’)** vraca koliko B zauzima string

Another difference between LEN and DATALENGTH is that the former excludes trailing spaces but the latter doesn’t.

### CHARINDEX

**CHARINDEX(substring, string[, start\_pos])** vraca poziciju prvog occurancea substringa unutar stringa.

SELECT CHARINDEX(' ', 'Itzik Ben-Gan'); => 6

### PATINDEX

PATINDEX(pattern, string) vraca prvi occurance patterna u stringu

Pattern koristi slican izraz poput patterna u LIKE-u

SELECT PATINDEX('%[0-9]%', 'abcd123efgh'); => 5; vrati index prvog broja

### REPLACE

**REPLACE(string, substr1, substr2)** zamenjuje sve occurances substring1 sa substring 2

SELECT REPLACE('1-a 2-b', '-', ':'); This code returns the output '1:a :b'.

Sa REPLACE mozes prebrojavati i broj occurancesa nekog karaktera. Na primer karakter koji trazis(zelis da prebrojavas) mozes da zamenis empty stringom. Izracunas LEN rezultata i oduzmes ga od LEN izvornog stringa.

SELECT (LEN(src\_strng) – LEN(REPLACE(src\_strng, ‘a’,’’)))

### REPLICATE

Replicate function replicates a string “x” number of times

REPLICATE(string, n)  
REPLICATE(‘abc’,3) => abcabcabc

Na primer combo sa RIGHT funkcijom   
**SELECT supplierid,   
 RIGHT(REPLICATE('0', 9) + CAST(supplierid AS VARCHAR(10)), 10) AS strsupplierid   
FROM Production.Suppliers;**

A close-up of a number

AI-generated content may be incorrect.

### STUFF

Ova funkcija sluzi da se substring unutar stringa zameni drugim substringom.

STUFF(string, pos, delete\_length, insert\_string)

Na primer:  
STUFF(‘xyz’, 2, 1, ‘kita’) => xkitaz

### UPPER & LOWER

Za uppercase i lowercase

### LTRIM, RTRIM & TRIM

Dozvoljava ti da removeujes leading/trailing ili oba leading & trailing characters.

To znaci removeovanje spaceova :)

LTRIM(string)

### FORMAT

Sluzi za formatiranje inputa na osnovu msft .net format stringa i culture-a.

**FORMAT(input , format\_string, culture)**FORMAT(10, ‘d10’) => 0000000010

FORMAT je skupa funkcija i ne bi je trebalo koristiti

### COMPRESS & DECOMPRESS

Kompresuju i dekompresuju po GZIP algoritmu  
COMPRESS(string)  
DECOMPRESS(string)

Prihvata character ili binary string kao input i vraca VARBINARY(MAX) typed value

Recimo imas stored procedure i zelis da storeujes CV-eve kandidata u bazi koji su kompresovani

INSERT INTO dbo.EmployeeCVs( empid, cv ) VALUES( @empid, COMPRESS(@cv) );  
Ovaj @cv input je NVARCHAR(MAX)  
Ovaj @empid je input parametar za employee

Kolona “cv” u dbo.EmployeeCVs ce biti tipa VARBINARY(MAX)

**DECOMPRESS** prihvata binary string kao input i vraca VARBINARY(MAX) kao output. Da bi dobio originalnu vrednost nazad, moraces castovati taj varbinary u ono sto ocekujes:  
  
**SELECT CAST(DECOMPRESS(COMPRESS(N'This is my cv. Imagine it was much longer.')) AS NVARCHAR(MAX));**

Pa na primer za onu kolonu gore bi taj decompress izgledao ovako:

SELECT empid,   
 CAST(DECOMPRESS(cv) AS NVARCHAR(MAX)) AS cv   
FROM dbo.EmployeeCVs;

### STRING\_SPLIT

Separateuje string u listu individualnih vrednosti

SELECT value FROM STRING\_SPLIT(string, separator[, enable\_ordinal]);

Za razliku od svih dosadasnjih funkcija, ovo je prva table function.   
**SELECT CAST(value AS INT) AS myvalue   
FROM STRING\_SPLIT('10248,10249,10250', ',')**

Ordinal flag moze ako imas mssql 2022 or later.

**SELECT CAST(value AS INT) AS myvalue, ordinal  
FROM STRING\_SPLIT('10248,10249,10250', ‘,’ , 1)**

Ovo je potom rezultat

A close-up of a number

AI-generated content may be incorrect.

### STRING\_AGG

Aggregate funkcija agregira listu stringova u jedan   
STRING\_AGG(input , separator) [WITHIN GROUP(order\_specification) ]

U sustini u toj target grupi funkcija konkatonira vrednosti input argumenta i razdvaja ih na osnovu separatora.

Da bi radio garanciju nad redosledom vrednosti, onda se koristi WITHIN GROUP

Naredni query agregira order ID-eve svakog customera rasporedjenih po najskorijem i razdvojeni zarezom

SELECT custid,   
 STRING\_AGG(CAST(orderid AS VARCHAR(10)), ',')   
 WITHIN GROUP(ORDER BY orderdate DESC, orderid DESC) AS custorders   
FROM Sales.Orders   
GROUP BY custid

A screenshot of a computer code

AI-generated content may be incorrect.

NOTE:  
Output STRING\_AGG je **nvarhar** osim ukoliko input nije **varchar.** U tom slucaju je i output **varchar**.

### LIKE Predicate

Koristi se da se isproverava da li character string matchuje specified pattern.

Characters:   
% => The percent(any size) wildcard  
\_ => The underscore(single character) wildcard  
[ABZ] =>The [<list of characters>] wildcard  
[A-Z] =>The [<character-character>] wildcard  
[A^B-Z] => The [^<character list or range>] wildcard

## Working with date and time data

Ovo ludilo podrzava x6 date and time formata.

DATETIME & SMALLDATETIME => Kao legacy formati  
DATE, TIME, DATETIME2 & DATETIMEOFFSET

A white sheet with black text

AI-generated content may be incorrect.

Storage poslednja x3 zavisi od preciznosti koju odaberes

### Literals

Kada treba da definises konstantu za date & time trebas da razmislis o nekoliko stvari.   
Naime T-SQL nema bas prirodan nacin za definisanje Date & Time konstanti pa moras definisati konstantu jednog tipa i potom je konvertovati u date & time.

Postoje i implicit konverzije, ali je kompletan fuckup to razumeti.   
**SELECT orderid, custid, empid, orderdate   
FROM Sales.Orders   
WHERE orderdate = '20220212';**

* Naime ovaj gore query je validan :) Ovaj dole ce biti uz cast:

**SELECT orderid, custid, empid, orderdate   
FROM Sales.Orders   
WHERE orderdate = CAST('20220212' AS DATE);**

A screenshot of a data form

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Nmg vise o ovome :)

### Filtering Date Ranges

Ako moras da filtriras po mesecu ili po godini, ima smisla koristiti YEAR i MONTH funkcije

SELECT \* FROM x WHERE YEAR(orderdate) = 2021  
SELECT \* FROM x WHERE MONTH(orderdate) = 2

ALI, vidis dragi moj naivni citaoce. Ovo sjebava indeksiranje :)   
Ne znam kako ni zasto, ali u sustini ne moze se efikasno iskoristiti pa bi se trebalo pridrzavati obicnih logickih operacija :)   
**Month varijanta: =>-----------  
SELECT orderid, custid, empid, orderdate   
FROM Sales.Orders   
WHERE orderdate >= '20220201' AND orderdate < '20220301';**

**YEAR varijanta: =>-----------  
SELECT orderid, custid, empid, orderdate   
FROM Sales.Orders   
WHERE orderdate >= '20210101' AND orderdate < '20220101';**

## Date and time functions

GETDATE, CURRENT\_TIMESTAMP, GETUTCDATE, SYSDATETIME, SYSUTCDATETIME, SYSDATETIMEOFFSET, CAST, CONVERT, PARSE, SWITCHOFFSET, TODATETIMEOFFSET, AT TIME ZONE, DATEADD, DATEDIFF and DATEDIFF\_BIG, DATEPART, YEAR, MONTH, DAY, DATENAME, DATETRUNC, DATE\_BUCKET, ISDATE, various FROMPARTS functions, EOMONTH, and GENERATE\_SERIES.

### GETDATE, CURRENT\_TIMESTAMP, GETUTCDATE, SYSDATETIME, SYSUTCDATETIME, SYSDATETIMEOFFSET

A screenshot of a computer

AI-generated content may be incorrect.

### CAST, CONVERT, PARSE & their TRY\_Counterparts

Sluze za konvertovanje input value-a u neki drugi target type. TRY\_verzija ovih funkcija ako konverzija failuje ce vratiti NULL umesto da failuje.

CAST(value as target\_data\_type) => SELECT CAST(SYSDATETIME() AS DATE)  
CONVERT(target\_data\_type, value, [style\_number])  
PARSE(value as target\_data\_type [USING culture])

TRY\_CAST, TRY\_CONVERT, TRY\_PARSE isto sve.

Treci opcioni argument je za definisanje npr stila konverzije tj npr kad iz char -> datetime pa hoces da definises u kom formatu si pisao taj datetime DD/MM/YYYY umesto YYYY-MM-DD na primer.

### SWITCHOFFSET Function

Podesava vrednost DATETIMEOFFSET-a u specificirani target offset “from UTC”.

SWITCHOFFSET(datetimeoffset\_value, UTC\_offset)

Umm SELECT SWITCHOFFSET(SYSDATETIMEOFFSET(), '-05:00');

So if the current system datetimeoffset value is February 12, 2022 10:00:00.0000000 –08:00, this code returns the value February 12, 2022 13:00:00.0000000 –05:00

SELECT SWITCHOFFSET(SYSDATETIMEOFFSET(), '+00:00');

Assuming the aforementioned current datetimeoffset value, this code returns the value February 12, 2022 18:00:00.0000000 +00:00.

### TODATETIMEOFFSET Function

Kreira DATETIMEOFFSET typed value iz lokalnog date and time value i offset-a iz UTC  
  
**TODATETIMEOFFSET(local\_date\_and\_time\_value, UTC\_offset)**

You will typically use this function when migrating non-offset-aware data to offset-aware data.

I dalje mislim da ne kontam ovo.

### The AT TIME ZONE Function

Uzima date and time value i konvertuje ga u **datetimeoffset** type za taj specified time zone.

**date\_time\_value AT TIME ZONE time\_zone**

Supported inputs: DATETIME, SMALLDATETIME, DATETIME2, and DATETIMEOFFSET

--- ima jos dosta o ovome, nadam se da mi nikada nece trebati.

### The DATEADD function

**DATEADD(part, n, dt\_val)**

Dodaje “jedinice”(n) na specificirani date part u date time value  
Validne vrednosti za part su:  
*year, quarter, month, dayofyear, day, week, weekday, hour, minute, second, millisecond, microsecond, and nanosecond*

DATEADD(year, 1, ‘20220212') => 2023-02-12 00:00:00.000

### DATEDIFF & DATEDIFF\_BIG Functions

Vracaju razliku izmedju x2 date and times na osnovu specified date part-a.  
DATEDIFF vraca rezultat u INT  
DATEDIFF\_BIG vraca rezultat u BIGINT type-u

DATEDIFF(date\_part, date\_1, date\_2) :INT  
DATEDIFF\_BIG(date\_part, date\_1, date\_2) :BIGINT

**SELECT DATEDIFF(day, '20210212', '20220212');** => 365

Vrati ti razliku u danima, jelte.

### DATEPART Function

Vraca integer koji predstavlja specified date\_part

DATEPART(date\_part, date) :INT

DATEPART(month, ‘20220221) => 2

Mozes koristiti i *day, weekday, dayofyear* kao part parametre.

### YEAR, MONTH & DAY Functions

Poticu iz DATEPART funkcije. Isti k samo sto jelte lakse za izrazavanje  
YEAR(date) == DATEPART(year, date)

### DATENAME Function

Vraca character string koji predstavlja “part” tog date\_value-a.

DATENAME(date\_value, part);

DATETNAME(‘20220201’, month) => ‘February’

Ukoliko ti je session language podesen na english dobices February. Ako ti je na italijanskom, onda dobijer “Februraroaroiujasfi” iliti sta god da je na italijanskom februar.

Ako requested part nema naziv(npr godina), onda samo dobijes broj koji je zapravo character :)

DATENAMT(‘20220201’, year) => ‘2022’

### DATETRUNC Function

DATETRUNC truncateuje tj. flooruje input date and time value do “pocetka specificiranog dela”

Ovo je predstavljeno u SQL SERVER 2022

**DATETRUNC(part, date\_value)**

Ako je date\_value tipa DATETIME, onda ce i output biti DATETIME. Ako je date\_value character string, onda ce output biti DATETIME2(7)

Valid part vrednosti: *year, quarter, month, dayofyear, day, week, iso\_week, hour, minute, second, millisecond, and microsecond*

DATETRUNC(month, ‘20120224’) => 2012-02-01 00:00:00.000

### ISDATE Function

Uzima character input i vraca 1/0(bool) ukoliko je character valid date ili nije

ISDATE(‘2022123493’) => 0

ISDATE(‘20220221’) => 1

### FROMPARTS Function

FROMPARTS funkcija prihvata integer input koji predstavljaju delove date & time vrednosti i konstruisu vrednost “requested tipa” iz tih delova

DATEFROMPARTS(year, month, day)  
DATETIME2FROMPARTS(year, month, day, hour, minute, seconds, fractions, precision)  
DATETIMEFROMPARTS(year, month, day, hour, minute, seconds, milliseconds)  
DATETIMEOFFSETFROMPARTS(year, month, day, hour, minute, seconds, fractions, hour\_offset, minute\_offset, precision)  
SMALLDATETIMEFROMPARTS(year, month, day, hour, minute)  
TIMEFROMPARTS(hour, minute, seconds, fractions, precision)

A number and numbers on a white background

AI-generated content may be incorrect.

### EOMONTH Function

Prihvata datum(date and time value) i vraca u DATE formatu respective end of month date

EOMONTH(input [, months\_to\_add])

**SELECT EOMONTH(SYSDATETIME())**

### The GENERATE\_SERIES function

Ovo je table funkcija koja vraca sekvencu brojeva u zahtevanom range-u.  
Introduced in MSSQL 2022

GENERATE\_SERIES(start\_value, stop\_value [, step\_value])  
Specificiras start i stop values i optionally step. Mozes -1 za step kako bi dobio decresing range, to probaj kad budes zapisivao u svesku.

SELECT value  
FROM GENERATE\_SERIES(1, 10)

A white sheet with black text

AI-generated content may be incorrect.

Mozes se malo igrati pa konvertovati u dane npr:   
A close-up of a date

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## Querying metadata

SQL Server ti nudi alate za queryovanje metadata objkata poput informacija u vezi tabela u bezi i kolona u tabeli.

To su catalog views, information schema views i system stored procedures & functions.

### Catalog views

Catalog views nude detaljisane informacije o objektima u bazi podataka, ukljucujuci i sql server specific data.

### SYS.TABLES VIEW

Npr zelis da izlistas tabele u bezi podataka ukljucujuci i njihove schema.  
To mozes kroz queryovanje ***sys.tables*** view-a

**SELECT SCHEMA\_NAME(schema\_id) as table\_schema\_name  
 , name as table\_name   
FROM sys.tables;**

SCHEMA\_NAME konvertuje schema ID integer u njegov naziv.

A table of sales

AI-generated content may be incorrect.

### SYS.COLUMNS Table

Da dobijes informacije o kolonama u tabeli.

SELECT name AS column\_name  
 , TYPE\_NAME(system\_type\_id) AS column\_type  
 , max\_length  
 , collation\_name  
 , is\_nullable  
FROM sys.columns WHERE object\_id = OBJECT\_ID(N’Sales.Orders’)  
A screenshot of a computer

AI-generated content may be incorrect.

### Information schema views

SQL Server supports skup viewova koji se nalaze u schema “INFORMATION\_SCHEMA”

Npr:  
***INFORMATION\_SCHEMA.TABLES*** view pokazuje base tabele u bazi zajedno sa njihovim schema names

SELECT TABLE\_SCHEHMA, TABLE\_NAME  
FROM INROMATION\_SCHEMA.TABLES  
WHERE TABLE\_TYPE = N’BASE TABLE’;

***INFORMATION\_SCHEMA.COLUMNS*** nudi vecinu dostupnih informacija o kolonama:

SELECT COLUMN\_NAME, DATA\_TYPE, CHARACTER\_MAXIMUM\_LENGTH, COLLATION\_NAME, IS\_NULLABLE   
FROM INFORMATION\_SCHEMA.COLUMNS   
WHERE TABLE\_SCHEMA = N'Sales' AND TABLE\_NAME = N'Orders';

## System stored procedures & functions

Internally one prave upite nad system catalog i vracaju malo razumnije medatata informacije.

Primeri:  
**EXEC sp\_tables** -> vraca listu objekata poput tabela i viewova koji moogu biti queryovani u trenutnoj bazi podataka

**EXEC sys.sp\_help @objname = ‘Sales.Orders’** -> vraca multiple result set sa osnovnim informacijama o objektu i info o kolonama, indexima, constraints & vise za neku tabelu.

**EXEC sys.sp\_columns   
 @table\_name = ‘Orders’  
 @table\_owner = ‘Sales’** -> info o kolonama u objektu

**EXEC sys.sp\_helpconstraint @objname = N’Sales.Orders’** -> info o constraintovima u objektu.

Potom imamo SERVERPROPERTY funkciju koja vraca requested property trenutne server instance

**SELECT SERVERPROPERTY(‘Collation’);** -> Vraca collation trenutne instance

**SELECT DATABASEPROPERTYEX(N’TSQLV6’, ‘Collation’)** -> Vraca property neke specificirane baze

**SELECT OBJECTPROPERTY(OBJECT\_ID(N’Sales.Orders’), ‘TableHasPrimaryKey’)** -> Vraca property specificiranog object name-a, npr ovo gore naznacava da li tabela ima primary key

**SELECT COLUMNPROPRTY(OBJECT\_ID(N’Sales.Orders’), N’shipcountry’, ‘AllowsNull’)**  
-> vraca requested property za kolonu. Npr ovaj gore otput govori o tome da li je kolona nullable ili nije.

## Exercises Sve do 116. strane

## JOINS

Unutar FROM clause-a table operatori mogu da vrse operacije nad tabelama.   
T-SQL supports x4 tipa table operatora:  
- Join (standard); ostali su non standard tj extenzije  
- Apply  
- Pivot  
- Unpivot

JOIN vrsi operacije nad x2 input tabele. Postoje cross joins, inner join i outer join.

### CROSS JOIN

Implementira samo x1 logical query processing fazu – a Cartesian Product.

U ovoj fazi radi se nad dveju tabela kao input i daje se Cartesian Product kao rezultat ove dve(Kartezijski proizvod?). U prevodu, svaki red iz jednog inputa je matchovan sa svakim rowom iz drugog. Matematicki: tabela\_1 ima *m* redova, a tabela\_2 ima *n* redova.

Rezultat je *m* x *n*

Postoje x2 sintakse -> SQL-92(recommended) & SQL-89

**SQL-92 sintaksa:**   
SELECT c.custId, e.empId  
FROM Sales.Customers as c  
 CROSS JOIN HR.Employees as e

**SQL-89 sintaksa:**SELECT c.custId, e.empId  
FROM Saleas.Customers as c, HR.Employees as e

SQL-89 sintaksa jeste “kraca”, ali znantno necitljivija. Nema performantnih razlika.

### INNER JOIN

Inner join sadrzi x2 faze procesiranja – Cartesian product izmedju x2 input tabele kao u cross join i onda filtrira redove na osnovu predikata kog specificiras.

Naravno, postoje SQL-89 & SQL-92 standard.

**SQL-92**

SELECT e.empid, e.firstname, e.lastname, o.orderid  
FROM HR.Employees as e  
 INNER JOIN Sales.Orders as o  
 ON e.empid = o.empid

U sustini matchuje se svaki employee row sa svakim order row-om koji imaju iste employee id’s.

Sta se zapravo desava :)   
Znaci prvo ide jedan cross join => 9 empoyee rows x 830 order rows = 7470 rows  
Potom join filtrira rows na osnovu “ON” predicate-a e.empid = o.empid sto vraca recimo 830 redova.

ON clause vraca samo redove za koje predicate evaluira na “TRUE”, dakle False i Unknown se odbacuju.

**SQL-89**

SELECT \*   
FROM HR.Employees as e, Sales.Orders as o  
WHERE e.empid = o.empid

No performance differences, interpreted as same shit. Jer u sustini desava se cross join pa onda filter uz “ON”. Sad mi je sve jasno.

### BEGIN Joins with extra steps(more join examples)

### Composite join

Composite join je join koji radi filter fazu inner joina na “vise uslova”  
Dakle nece biti samo npr ON e.empid = o.empid nego ce biti

ON e.empid = o.empid AND o.propNeki = e.propNeki

Dakle, taj “ON” je kao “WHERE” filter, mozes drugacije stvari da redjas tu.

SELECT OD.orderid, OD.productid, OD.qty, ODA.dt, ODA.loginname, ODA.oldval, ODA.newval   
FROM Sales.OrderDetails AS OD   
 INNER JOIN Sales.OrderDetailsAudit AS ODA   
 ON OD.orderid = ODA.orderid AND OD.productid = ODA.productid   
WHERE ODA.columnname = N'qty';

### Non-Equi Joins

Kada JOIN condition ukljucuje equality operator, kaze se da je to “equi join”. Kada join condition ukljucuje bilo koji operator sem equality(jednakosti), onda je to non-equi join.

Standard SQL supportuje “natural join” feature koji radi join izmedju x2 tabele na osnovu kolona sa identicnim nazivima. Na primer T1 NATURAL JOIN T2 joinuje redove na osnovu kolona koje imaju iste nazive sa obe strane. T-SQL nema implementaciju natuarl joina.

Npr za kreiranje unique pairs of employees:

SELECT e1.empid, e2.empid  
FROM HR.Employees as e1 INNER JOIN HR.Employees as e2  
ON e1.empid < e2.empid

Na primer ovo ce jointovati sve employees koji nemaju isti employee id. Da je samo cross join bio koristen, onda bi rezultat ukljucio i self pairs(1 sa 1) i mirror pairs(1 sa 2 i 2 sa 1).

### Multi join queries

JOIN operator radi na x2 tabele, a kada se pojavi vise tabela onda ce se JOIN izvrsavati redosledom kojim su joinovi napisani.

SELECT \*   
FROM Customers c INNER JOIN Orders o ON c.custid = o.custid  
 INNER JOIN OrderDetails od ON o.orderid = od.orderid

Ovde gore ce se desiti sledece:  
1. inner join izmedju customers i orders  
2. inner join izmedju orderdetails i rezultata prethodnog joina

### END Joins with extra steps(more join examples)

### OUTER JOINS

Outer joins imaju samo x1 standard syntax – onaj sa JOIN & ON clauseom. Postji samo SQL-92 verzija.

Outer join ima x3 faze. Cartesian product, ON filter & treca faza zvana Adding outer rows koja je unique samo za ovaj tip joina.

U outer join ti obelezavas tabelu kao “preserved table” koriscenjem sledecih keywords:  
- LEFT OUTER JOIN  
- RIGHT OUTER JOIN  
- FULL OUTER JOIN

“OUTER” je optional. LEFT keyword znaci da ce rows leve tabele biti zadrzani. RIGHT keyword znaci da ce redovi desne tabele biti zadrzani.   
FULL OUTER JOIN znaci da su redovi u obe tabele perserved.

Ova treca logicka faza outer joina koja se naziva adding outer rows identifikuje redove iz perserved tabele koji nisu pronasli matcheve u ON predicate-u. Potom ih dodaje u result table i koristi “NULL” kao placeholder za atribute iz nonpreserved strane u tim perserved redovima.

SELECT c.custid, c.companyname, o.orderid  
FROM Sales.Customer c  
LEFT OUTER JOIN Sales.Orders o   
 ON c.custid = o.custid

In the query above, all customers will be matched to their respective orders. Those which haven’t made any orders will also be shown, but their orderid will be NULL.

U slucaju da nisi znao(znao si, ali ako zaglupis odjednom) – Conceptually prvo se izvrsava FROM clause i joinovi u njemu. Ukoliko na tvoj query dodas i WHERE clause, taj WHERE ce se izvrsiti nakon sto su svi table operatori izvrseni i on ce filtrirati samo end result.

### Including missing values

Na primer imas scenario u kom zelis da filtriras sve porudzbine od 1. jan 2020 do 31. dec 2022. U tim datumima ne zelis samo da vratis dane kada si imao porudzbine nego i dane kada nisi imao porudzbine, znaci kada i orderdate-a nema.

Da bi se solveovao problem, mozes napisati query koji vraca sekvencu svih datuma u requested periodu i onda odradis outer left join izmedju tog skupa i orders tabele.

Da bi dobio sekvencu datuma, ono sto je lik uradio jeste napravio tabelu sa brojevima – sekvence npr.

SELECT DATEADD(day, n-1, CAST(‘20200101’ as DATE)) as orderdate  
FROM dbo.Nums  
WHERE n <= DATEDIFF(day, ‘20200101’, ‘20221231’)+1  
ORDERBY orderdate

Potom ces doci i extendovati taj query sa orderima

SELECT DATEADD(day, n-1, CAST(‘20200101’ as DATE)) as orderdate,   
 o.orderid, o.custid, o.empid  
FROM dbo.Nums   
 LEFT OUTER JOIN Sales.Orders o   
 ON o.orderdate = DATEADD(day, n-1,CAST(‘20200101’ as DATE))  
WHERE n <= DATEDIFF(day, ‘20200101’, ‘20221231’)+1  
ORDERBY orderdate

dbo.Nums je u sustini tabela sa brojevima 1-1000 npr. Od SQL Server 22 mozes koristiti GENERATE\_SERIES funkciju umesto ovog workarounda.

### Using outer joins in a multi-join query

Dakle, znamo da table operatori se izvrsavaju redosledom kojim su napisani. Zato pazi koju stranu perserveujes. Takodje imaj na umu da outer rows za atribute iz nonperserved strane imaju “null” i samim tim UNKNOWN se odbacuje u ON filteru.

### Filtering attributes from the nonperserved side of an outer join

Uzeti u obzir sledeci query:

SELECT \*   
FROM Sales.Customers c  
 LEFT OUTER JOIN Sales.Orders o ON o.custid = c.custid  
WHERE o.orderdate >= ‘20220101’;

Dakle ovde se prvo vracaju svi customers sa njihovim porudzbinama, ali onda posto cemo imati left join gde se vrate i customeri bez porudzbina, a mi ukjucujemo WHERE filter na osnovu vrednosti koja ce biti NULL za customere koji nemaju porudzbine, mi cemo u sustini imati kao neki rezultat iz INNER JOIN-a. Oni sa NULL orderdate-om ce rezultovati u UNKNOWN. Ovo je cest problem, bolje razmisljaj :)

### Using COUNT aggregate with outer joins

Drugi cesti bug jeste koriscenjem COUNT sa outer joinovima. Kada gripises rezultat outer joina i koristis COUNT(\*) aggregate, taj agregate ukljucuje obe inner rows i outer rows jer prebrojava redove bez obzira na njihov sadrzaj.

Uglavnom ne bi trabalo da ukljucujes outer rows into consideration u svrhe brojenja.

Na primer treba da prebrojavas broj porudzbina svakog customera

SELECT c.custid, COUNT(\*) as numoforders  
FROM sales.customers c  
 LEFT OUTER JOIN sales.orders o ON c.custid = o.custid  
GROUP BY c.custid

Customeri 22 i 57 koji nemaju ordere ce imati outer row u rezultati joina pa samim tim ce im numoforders reci “1”.

Zato umesto COUNT(\*) ides COUNT(o.orderid) jer COUNT(NULL) je ignorisan, a outer row to nece imati.

## Exercises sve do 149. stranice

## Subqueries

SQL Supportuje pisanje query-a unutar query-a tj nesting queries.

“Najspoljnii”(outer most) query je taj koji ce vratiti krajnji rezultat. Inner query se evaluira u run time i npr rezultat moze da se menja u zavisnosti od podataka, jelte. Koristi se umesto da sad hardkodiras neku konstantu.

Sub-query moze biti:  
- self contained  
- correlated

Self contained subquery nema zavisnosti od tabela spoljnog query-a. Correlated ima dependency tabela spoljnog querya.

Subquery result moze biti single valued, multivalued ili table valued.

Ovaj chapter se vise fokusira na single valued subqueries(scalar queries) i multivalued subqueries.

### Self contained subqueries

Self contained subqueries su nezavisni query-i od outer querya.

Logicki redosled izvrsavanja je takav da se prvo izvrsava subquery kod, a tek onda outer query. Potom outer query koristi rezultat subquery-a.

Na primer zelis da uzmes najskuplji proizvod i zapamtis ga.

DECLARE @priciestItem INT = (SELECT MAX(price) FROM Products)

ili umesto da ga pamtis

SELECT \* FROM Products WHERE price = (SELECT MAX(price) FROM Products)

Scalar subquery mora da returnuje single value. Ako returnuje vise vrednosti onda ce failovati u toku runtime.

Zasto ce failovati ako ne vrati skalarni rezultat? Pa jer equality je binarni operator koji poredi jednu vrednost sa leve sa jednom vrednostii s desne strane. Da se vrati vise vrednosti, puca.

Ukoliko subquery ne vrati nikakvu vrednost, onda se taj prazan rezultat konvertuje u **NULL!**

Poredjenej sa NULL yielduje u UNKNOWN i taj query filter potom nece vratiti rezultat jer samo oni koji evaluiraju na TRUE vracaju.

### Self contained multivalued subquery

Isti k samo ces raditi sa npr **IN** predicateom

SELECT \* FROM Products   
WHERE productId IN (SELECT productId FROM Products WHERE price > 200)

Subquery nekad moze da ima bolji performans od joina. Nekada join moze imati bolji performans. Underlying DB engine ce svakako dodati “optimizacione tehnike” pa ne mozes uvek 100% biti siguran.

Pristup koji ima covek koji je napisao knjigu je sledeci: Prvo napisi intuitivni query koji daje tacan rezultat. Potom podesavaj performans :)

Poslednji primer:  
Treba da napises query koji vraca sve incomplete orderid’s koji fale izmedju minimum i maximum ones u tabelama.

SELECT n FROM dbo.Nums   
WHERE n BETWEEN   
 (SELECT MAX(orderid) FROM Orders) AND   
 (SELECT MAX(orderid) FROM Orders)  
AND NOT IN (SELECT orderid FROM CompletedOrders)

### Correlated subqueries

Correlated subqueries su subqueris koji referuju atributima spoljnih query-a u kojima su nestovani. To znaci da je subquery zavistan od outer query i ne moze biti invoked kao standalone query.

Logicki gledano, subquery se evaluira separately za svaki outer row u logical query processing step u kom se pojavljuje.

Primer: kod koji vraca najskuplje order za customera(probaj i sa group by)

SELECT o1.custid, o1.orderid  
FROM Orders o1  
WHERE orderid = (SELECT MAX(o2.orderid) FROM Orders o2 WHERE o2.custid = o1.custid)

TLDR: Subquery gets executed for each row in the outer query.

E sada, mnogo je teze troubleshootovati ovakve queryje. Nazalost. Ne mozes samo doci i executeovati subquery jer vrednost zavisi od outer query. Zato sta se radi je sledece:  
Za testing subquery, uzmes jednu vrednost(red) iz outer query-a i hardkoridas u inner query.

Na primer: SELECT MAX(o2.orderid) FROM Orders o2 WHERE o2.custid = 86

Za red sa customerid gde je 86 mozes da vidis sta ce subuery vratiti

Potom uzmes taj rezultat i testiras outer query

SELECT o1.custid, o1.orderid  
FROM Orders o1  
WHERE orderid = 102534

Mozes tako da se igras, malo je zeznutno, ali funkcionise.

### EXISTS Predicate

EXISTS predicate prihvata subquery kao input i vraca TRUE ako subquery vrati bilo kakav row. Rezultat bez rows vraca FALSE

SELECT custid, companyName  
FROM Customers c  
WHERE coutnry = ‘Spain’  
 AND EXISTS(SELECT \* FROM Orders o WHERE o.custid = c.custid)

Query iznad vraca customere iz spanije koji su napravili neku kupovinu.

Mozes koristiti i NOT EXISTS koji ce vratiti “true” ukoliko nema redova koji postuju uslov, jelte.

Na primer za ovaj query iznad da si koristio NOT EXISTS onda bi dobio spance koji nemaju porudzbine.

PERFORMANCE EXISTS QUERY-a; Dakle DB engine shvata da je dovoljno vratiti jedan rezultat za EXISTS query. Shvati to kao short-circuit evaluation. Same applies na IN Predicate. Cak i za (\*) koje je smatrano losom praksom, EXISTS je okej. Predicateu je stalo samo do postojanja matching redova bez obzira sta napises u SELECT tako da ce db engine ignorisati subquery-s select list.

To ne znaci da (\*) nece biti resolveovan u EXISTS(SELECT \* FROM...), ali performance differene sa necim poput EXISTS(SELECT 1 FROM...) je bukvalno neprimetno.

Takodje EXISTS koristi two valued logic, a ne three valued logic.

### Returning previous or next values

Recimo da moras queryati Orders table i vratiti za svaki order informacije o trenutnom orderu i prethodni orderid. Taj deo “previous” implicira redosled kog tabele nemaju, jelte.

Jedan pristup ovome jeste “maksimalna vrednost koja je manja od trenutne vrednosti.

SELECT orderid, orderdate, empid, custid,  
 (SELECT MAX(o2.orderid)   
 FROM Orders o2 WHERE o2.orderid < o1.orderid) as prevorderid  
FROM Orders o1

### Using running aggregates

Running aggregates su agregati koji akumuliraju vrednosti na osnovu nekog redosleda.

A close-up of numbers

AI-generated content may be incorrect.  
Recimo da moras da izracunas zbir svih “quantity” vrednosti do neke goine.   
Za 2021 to ce biti quantity iz 2020 i 2021. Za 2020 samo iz 2020.

SELECT orderyear, qty,   
(SELECT SUM(o2.qty) FROM OrderTotalsByYear o2   
WHERE o2.orderyear < o1.orderyear) as runqty  
FROM OrderTotalsByYear o1  
ORDER BY orderyear

* Ne kontam sta je fora sa ovim. Ovo je obican correlated query gde se taj correlated query izvrsava u SELECT(skoro poslednjoj fazi) clauseu

### BEGIN - Dealing with misbehaving subqueries

### Null Trouble

Uzmi query koji ce vratiti customere koji nemaju ordere:  
SELECT \* FROM Customers c WHERE custid NOT IN (SELECT o.custid FROM Orders o)

I dobijes neki rezultat, set od x2 row-a.

A close-up of a list of words

AI-generated content may be incorrect.

E sada, odradi insert u orders tabelu gde ce jedan custid imati vrednost “NULL”  
Rezultat ce biti prazan skup, tj nista neces dobiti. Kako to.

Jer jebeni three valued logic.

Sta se desava?  
Posto subquery vraca multiresult u taj “IN” nad kojim imas negaciju sa “NOT”, desi se da ukoliko se pojavi NULL u customer id orders, ne mozes sa preciznoscu reci da li jeste ili nije taj customer id tu. Jer NULL predstavlja odsustvo vrednosti, jelte.

Da imas non null multivalue u IN onda mozes sa sigunoscu reci da li je neki custid u tom spisku ili nije.

IN predicate returnuje UNKNOWN za customera koji nije napravio order(custid: 22) jer ce se porediti sa NULL u jednom momentu i dobices UNKNOWN. FALSE ili UNKNOWN yielduje u UNKNOWN.

Kada se to prevede u “vrednost” ti taj NOT IN(1, 2, ..., NULL) u momentu dobijes NOT UNKNOW sto je => UNKNOWN

TLDR => NULL u “IN” operatoru ce uvek rezultirati u UNKNOWN

Resenje:  
NOT EXISTS (SELECT \* FROM Orders o WHERE o.custid = c.custid)

ili

NOT IN (SELECT o.custid FROM Orders o WHERE o.custid IS NOT NULL)

### Substitution errors in subquery column names

Lik je dosao i rekao “e pazi na typo”

### END – Dealing with misbehaving subqueries

### Exercises sve do 177. stranice

## Derived Tables

Derived tabele su definisane u FROM clause u outer query-u. Njihov scope postojanja jeste outer query tako da cim se outer query izvrsi, one prestaju da postoje.

SELECT \*   
FROM (  
 Select custid, companyname   
 from sales.customers where country = ‘usa’) as USACustomers  
=> specificiras query koji predstavlja tabelu, posle toga dodas “AS” gde kazes derived table name.

**Da bi expression bio ispravan, query mora ispuniti sledece zahteve:**  
- Da nije u SNS  
- Redosled row-ova da nije zagarantovan => ORDER BY je dozvoljen samo ako ne sluzi za prezentacionu svrhu. Znaci ako koristis OFFSET-FETCH npr.   
- Sve kolone moraju imati nazive  
- Svi nazivi kolona moraju biti unique

### Assigning column aliases

Recimo scenario u kom zelis da grupises po alias column name

SELECT YEAR(orderdate) as orderyear  
FROM Orders  
GROUP BY orderyear

Ovde ces dobiti error da orderyear ne postoji, jelte.

Resenje: derived table

SELECT \* FROM (  
 SELECT YEAR(orderdate) as orderyear  
 FROM Orders) as xyz  
ORDER BY orderyear

Nema performance benefita ni gubitaka. MSSQL ce u pozadini izvrsiti nesto poput sledeceg:

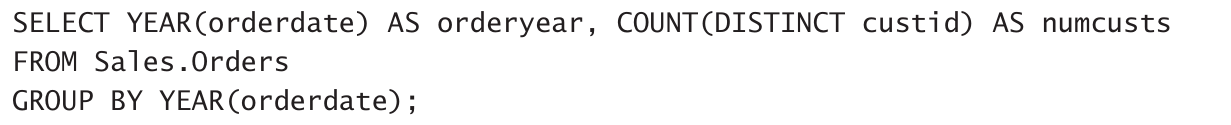


Table expressions se uglavnom koriste iz logickih razloga. Ne iz performantnih razloga.

### Using arguments

Argumenti, upoznat si sa njima. Dakle mogu biti lokalni argumenti ili argumenti neke rutine poput stored procedure ili funkcije.

DECLARE @var\_name AS var\_type = var\_value

DECLARE @custid AS INT = 5;  
SELECT \* FROM Orders WHERE custid = @custid;

AS je optional.

### Nesting

Mozes nestovati

SELECT \* FROM (  
 SELECT kita, kita2 FROM (  
 SELECT ckita, ckita2 FROM orders WHERE ckita = 22, ckita2 <55) as nest 2) as nest1

Problem, citljivost nije ni za kurac

### Multiple references

Jos jedan problematican aspekt derived tabela jeste u slucajevima kada trebas da joinujes vise instanci istih. JOIN tretira dva inputa kao skupovi gde jelte, skup nema order.

To znaci da ako definises deived table i aliasujes ga kao input joina, ne mozes referovati isti alias u drugi input joina.

Moras

SELECT \*  
FROM (inner\_query) as cur  
LEFT JOIN (that\_same\_inner\_query) as prv

### Common table expressions

Common table expressiosn CTEs su jos jedan standard form table expressiona slican deriviranim tabelama, ali sa nekoliko bitnih “advantages”(zaboravih rec).

Definisane su koriscenjem **WITH statementa**

**WITH <CTE\_NAME>[(target\_column\_list)]  
AS   
(<inner\_query\_defining\_CTE>)  
<outer\_query\_against\_CTE>**

Inner query mora pratiti sve isto kao i derived table expression.

WITH USACusts AS   
(  
 SELECT custid, companyname  
 FROM Sales.Customers  
 WHERE country = ‘USA’  
)  
SELECT \* FROM USACusts;

Kao i sa derived tabelama, cim se outer query zavrsi tako ce i ta tabela biti obrisana.

### Assigning column aliases in CTEs

CTEs podrzavaju dva vida column aliasinga. Inline i external.

Za inline formu, to se specificira unutar samog selecta u common expressionu.  
Sto se tice external formu, to se specificira kao lista kolona u zagradi odmah nakon CTE imena.

**Inline**:   
WITH C AS   
(  
SELECT YEAR(orderdate) as orderyear, custid FROM Sales.Orders  
)

**External**:  
WITH C(orderyear, custid) AS  
(  
SELECT YEAR(orderdate, custid FROM Sales.Orders  
)

### Koriscenje argumenata u CTE

Kao i sa derived tabelama mozes koristiti argumente u inner query da bi definisao CTE.

DECLARE @empid as int = 3;

WITH C AS  
(  
SELECT \* ....WHERE empid = @empid  
)...

### Defining multiple CTE

WITH C1 as  
(  
...  
),  
WITH C2 as  
(  
SELECT ... FROM C2 c2 GROUP BY...  
)  
SELECT \* FROM C1 join C2 on ....

Zasto je ovo kul? Pa jer mozes da referenciras vrednosti iz jednog CTE sa drugim CTE. Ne moras ih nestovati, samo ih razdvojis kroz razmak. Svaki CTE moze referencirati prethodni CTE i mozes queryovati sve definisane CTE’s.

### Multiple references in CTEs

Cinjenica da se prvo definise CTE ti dozvoljava da u outer query-u saljes upit ka vise instanci istog CTE-a

WITH C AS   
(...)  
SELECT \* FROM C as c1 JOIN C as c2 ON ...

### Recursive CTEs

CTEs imaju podrsku za rekurzije. Rekurzivni CTE su definisani by the SQL standard i definisan je na osnovu barem x2 query-a(more are possible) – gde je bar jedan query anchor member i bar jedan recursive member.

Basic rekurzija izgleda nekako ovako:  
WITH <CTE\_Name>[(<columns>)]   
AS   
(  
<anchor\_member>  
UNION ALL   
<recursive\_member>  
)

<outer\_query\_against\_CTE>

**Anchor member** je query koji vraca validni relacioni result table – kao query koji sluzi za definisanje nerekurzivnog table expressiona. On se invokeuje samo jednom.

**Recursive member** jer query koji ima reference ka CTE name i invoked je sve dok se ne vrati prazan skup. Referenca ka CTE name predstavlja prethodni result set.

Prvi put pri izvrsavanju rekurzivnog membera, prethodni result set predstavlja ono sto je anchor member vratio. Pri svakom sledecem izvrsavanju rekurzivnog clana, referenca na CTE predstavlja result set kog vrati prethodno izvrsenje rekurzivnog membera.

CTE name represents:  
1st invocation => anchor member result  
2nd invocation => recursive member result  
3rd invocation => recursive member result  
... => until emtpy set

Oba queryja moraju biti kompatibilna u smislu broja kolona i tipa podataka koje vracaju. Referenca na CTE name u outer query predstavlja unifikovani rezultat skupova invokacija anchor membera i SVIH invokacija recursive membera.

Recimo sada u praksi -> Zelis da vratis informacije o zaposlenom i svim njegovim subordinates na svim nivoima(direct ili indirect).

**WITH EmpsCTE AS   
(  
SELECT empid, mgrid, firstname, lastname  
FROM HR.Employees  
WHERE empid = 2**

**UNION ALL**

**SELECT C.empid, C.mgrid, C.firstname, C.lastname  
FROM EmpsCTE as P  
INNER JOIN HR.Employees AS C ON C.mgrid = P.empid  
)  
SELECT rmpid, mgrid, firstname, lastname  
FROM EmpsCTE;**

Anchor member queryuje za employee sa ID od 2.  
Recursive member uzima i joinuje od tog membera, sve druge zaposlene kojima je on menadzer. Potom svi oni kojima je on menadzer se uzimaju i gleda se kome su oni menadzer.

U slucaju logicke greske u JOIN predicateu u rekurziji ili ako ima problema sa podacima kojji rezultuju u “ciklusima”, recursive member moze biti invokeovan beskonacno puta(endless loop). Kao safety measure, sql server restricts broj rekurzivnih invokacija na 100 po defaultu.

Code ce failovati ako je recursive member invokeovan vise od 100 puta. Mozes default ponasanje promeniti sa OPTION(MAXRECURSION n) na kraj outer query-a gde je n integer od 0-32767.

Ako zelis da uklonis ogranicavanje izvrsenja rekurzivnog querya, onda stavi MAXRECURSION 0. SQL Server stores medjurezultate izvrsavanja izmedju anchor i rekurzive membera unutar tempdb. Ako removeujes restriction i imas runaway query, work table ce se poprilicno napuniti i query se nikada nece zavrsiti.

## Views

Derived statements & CTE’s imaju single statement scope i nisu reusable. Views & inline table-valued functions(Inline TVFs) su x2 tipa table expressiona cije se definicije cuvaju kao objekti u bazi pa su reusable.

U vecini slucajeva, views i TVFs se tretiraju kao derived tables i CTEs.

**CREATE OR ALTER VIEW Sales.USACusts  
AS   
SELECT custid, companyname, ...  
FROM Sales.Customers  
WHERE coutnry = ‘USA’;  
GO**

**GO command** se ovde koristi da bi terminateovala “batch” u T-SQL. Bice kasnije objasnjeno.

CREATE OR ALTER syntax kreira ili alteruje definiciju objekta. CREATE VIEW ce samo kreirati npr.

**SELECT \* FROM Sales.USACusts;**

### Views & Order by

Query koji koristis da bi definisao view mora ispostovati sve sto mora i inner query. Znaci bez orderinga, sve view kolone moraju imati ime i imena moraju biti unique.

Uz TOP ti mozes u sustini koristiti ORDER BY i ljudi misle da ces kreirati “ordered view”. Iako je code tehnicki ispravan i view je kreiran, treba da znas da ako outer query nad tvojim viewom nema ORDER BY clause, prezentacioni order nije zagarantovan.

Cak iako ti izgleda kao da je ordered, to mozda moze samo da ti se cini jer je takva pozicija podataka.

### View options

Kada kreiras ili alterujes view, mozes specificirati view attributes & options kao deo view definitiona.   
Unutar headera viewa pod WITH clause, mozes specificirati atribute poput:  
- ENCRIPTION  
- SCHEMABINDING

Na kraju querya mozes specificirati:  
- WITH CHECK OPTION

### ENCRYPTION Option

Ova opcija je dostpuna kada kreiras ili alterujes view, stored procedures, triggers & user defined functions.

Encryption indicates da SQL Server ce interno storeovati text u definiciji objekta u “obfuscated formatu”. Taj obfuscated text nije direktno vidljiv koristnicima kroz bilo koji catalog objkeata, samo privileged useri mogu da gledaju.

Recimo imas: **CREATE OR ALTER VIEW Sales.USACusts AS SELECT   custid, companyname, contactname, contacttitle, address,   city, region, postalcode, country, phone, fax FROM Sales.Customers WHERE country = N'USA';  
 GO**

Ukoliko pokusas da pribavis definiciju objekta na sledeci nacin:  
**SELECT OBJECT\_DEFINITION(OBJECT\_ID(‘Sales.USACusts’))**

Dobices rezultat. Da si postavio ENCRYPTION opciju, ne bi dobio output.

Da bi postavio opciju koja se postavlja u headeru onda  
**CREATE OR ALTER VIEW <view\_name> WITH ENCRYPTION  
AS   
SELECT \*  
...  
GO**

I ukoliko sada ponovo pokusas da dobijes definiciju objekta, onda dobijes “NULL”

Bilo koja alternativa object definition funkciji ce ti dati neki error ili nista.

### SCHEMABINDING Option

Dostupna je Viewovima, UDFs i natively compiled modulima; binduje schema referenced objekta i kolona u schemu referencing objekta.

Naznacava da referenced objects ne mogu biti dropped i da referenced collumns ne mogu biti dropped ili promenjene.

**CREATE OR ALTER VIEW <> WITH SCHEMABINDING  
AS   
SELECT \*, address,... FROM Sales.Customers WHERE <>  
GO**

I sada ako pokusas da droppujes **address** kolonu iz **Sales.Customers** dobijas error:

A white background with black text

AI-generated content may be incorrect.

Bez schemabindinga bi mogao lagano dropovati. Sa schemabindingom ces izbeci runtime greske u kojima ces queryovati kolone tabele koje ne postoje u slucaju izmene podataka.

Da bi se supportovao schemabinding, postoje pravila:

* Bez koriscenja \* u SELECT -> izlistaj sve column names
* moras koristiti schema qualified two part names

### CHECK Option

Sluzi kako bi sprecio modifikacije kroz view koje bi mogle uticati na inner query filter tog view-a. Npr query koji definise view USACusts filtrira customere na osnovu drzave da je United States.

Kada to uradis bez check-a, to znaci da mozes insert kroz view customera iz drugih zemalja i mozes update country postojecih customera kroz taj view.

Npr sledeci kod ce dodati customera iz US-a kroz view.

**INSERT INTO Sales.USACusts(companyname, contactname, ...)  
VALUES (‘customer abcde’,...), (‘customer\_2 abcde’,...)**

Row je insertovan kroz view unutar customers table.

E sada, posto si uneo x2 customera kroz view koji npr nisu imali US kao country nego recimo **UK**, onda kada budes izvrsio view neces ih dobiti nazad.

**SELECT custid, companyname, country  
FROM Sales.USACusts  
WHERE companyname LIKE ‘customer%’;**

Kako bi odradio prevenciju modifikacija koji se conflictuju sa view’s filter, dodajes WITH CHECK PTION na kraj querya koji definise view.

CREATE OR ALTER VIEW Sales.USACusts WITH SCHEMABINDING  
AS   
SELECT \*   
FROM Sales.Customers  
WHERE country = ‘USA’  
WITH CHECK OPTION;  
GO

I sada kada pokusas da insertujes nesto sto se conflictuje sa viewovim filterom, dobijas error.

### Inline Table-Valued Functions

Inline TVFs su reusable table expressions koje podrzavaju input parametre. TVF’s su slicne viewovima pa se nekad i referuju kao “views with inputs”, ali ne formalno.

T-SQL podrzava i multi-statement TVF koji populateuje i returnuje table variable. To nije table expression jer se ne bazira na queryu

**CREATE OR ALTER FUNCTION dbo.GetCustOrders  
 (@cid AS INT) RETURNS TABLE  
AS   
 RETURN   
SELECT orderid, shippeddate, custid, empid, orderdate,...  
FROM Sales.Orders  
WHERE custid = @cid;  
GO**

Inline TVF’s header ima mandatory RETURNS TABLE clase sto znaci da funkcija vraca table result. Takodje i mandatory RETURN clause pre inner query-a.

Tu je i input parametar.

**SELECT orderid, custid FROM dbo.GetCustOrders(1) as O**

MAKE SURE YOU PROVIDE AN ALIAS!

Nije uvek requirement, ali jako dobra praksa i imas manje sansi za errore.

Takodje mozes TVFs mogu biti naravno i deo JOIN-a.

DROP FUNCTION IF EXISTS dbo.GetCustOrders;

## The APPLY Operator

Apply operator je table operator koji se koristi u FROM clause-u querya. Postoje x2 podrzana tipa apply-a: CROSS APPLY & OUTER APPLY

APPLY performs its work in logical query phases kao i JOIN. CROSS APPLY implements samo x1 logical query processing phase dok OUTER APPLY implements x2 phases.

APPLY nije standard. Njegov SQL Standard counterpart se naziva **LATERAL**, ali standardna forma nije implementirana u SQL Serveru.

APPLY operator radi nad x2 tabele, left & right(nek ih tako nazovemo). Right table je cesto derived table ili TVF. CROSS APPLY primenjuje right table na svaki row leve tabele i kreira result table sa unifikovanim rezultatima.

Slican je cross joinu, istina i naredna x2 querya ce vratiti isti rezultat:

**SELECT S.shipperid, E.empid  
FROM Sales.Shippers AS S CROSS JOIN HR.Employees AS E;**

**SELECT S.shipperid, E.empid  
FROM Sales.Shippers AS S CROSS APPLY HR.Employees AS E;**

Join tretira x2 inputa kao skupove pa stoga ne postoji order izmedju njih. Sa APPLY leva strana se prva evaluira i potom se desna strana evaluira za svaki red leve tabele.

Right side moze imati reference elementima sa leve. Te reference su u ustini correlations.

Na primer da vratis x3 most orders za svaki customer:

SELECT \*   
FROM Sales.Customers AS C  
 CROSS APPLY (SELECT TOP (3) orderid FROM Sales.Orders AS O WHERE O.custid = C.id  
 ORDER BY orderdate DESC, orderid DESC) AS A

Ukoliko leva strana nema rezultata, onda krajnji rezultat ce biti “nista”. Znaci nema na cemu da se applyuje desna tabela. Ukoliko leva strana ima rezultata, ali nema na cemu da se matchuje tj derived table je empty set, onda takodje se nista ne returnuje.

Ukoliko na primer zelis da bez obzira na rezultat desne strane vratisrowove sa leve strane cak iako nema matchova, onda koristi OUTER APPLY.

Druga logical faza ovo operatora zadrzava sve leve row-ove.   
Ona zadrzava row-ove sa leve strane za koje nije bilo matcheva sa desne strane i koristi NULLs kao placeholdere. U sustini je skoro ekvivalnentno sa LEFT OUTER JOIN. Ne postoji ekvivalent za RIGHT OUTER JOIN zbog prirode APPLY-a.

Tako da npr kombinacija za TVFs i CROSS APPLY bi za top 3 rows bi izgledalo ovako:  
A screenshot of a computer code

AI-generated content may be incorrect.

### Exercises do 210. stranice

## SET Operators

Set operatori(operatori skupova) su operatori koji kombinuju redove iz dva querya koji davaju skupove kao rezultate. Neki operatora otklanjaju duplikate iz tih rezultata pa vracaju skup, dok drugi bas i ne pa vracaju multiset.

To su: UNION, UNION ALL, INTERSECT, EXCEPT.

**Input query1  
<set\_operator>  
Input Query2  
[ORDER BY ...];**

Kako set operator radi sa skupovima, tako rezultati query-a ne smeju biti ordered. Query sa order by vraca ordered result, a ne multiset.

Ovar orderby se primenjuje na sam rezultat set operatora, to smes.

Dva input querya moraju vratiti rezultat sa istim brojem kolona i svaka kolona mora da ima podudarajuce tipove podataka. To znaci “taj tip podataka” ili da moze da se implicitno konvertuje u higher data type. Uz cast/convert mozes konvertovati naravno.

Imena kolona krajnjeg rezultata zavise od 1. query-a tako da pises aliase u 1. query ako ti treba. Best practice je da sve kolone u svim query-ima imaju isti naziv.

Kada set operator poredi dva reda izmedju x2 inputa on koristi distinctness-based comparison, a ne equality base comparison. Taj distinctness-based comparison je isti kao u distinct predicate-u koji tretira NULL kao non-null value za comparison purposes gde jedan NULL se ne razlikuje od drugog NULL-a, a NULL se razlikuje od vrednosti broja 13 npr.

### UNION Operator

Unifikuje x2 input query-a. Ako se row pojavljuje u jednom skupu, pojavice se i u rezultatu.

T-SQL supports UNION ALL & UNION(implicit DISTINCT) flavors UNION operatora.

SELECT country FROM HR.Employees  
UNION  
SELECT country FROM HR.Employees

Ce vratiti ***m*** broj rezultata. DISTINCT se implicitno primenjuje i duplikati se otklanjaju.

### UNION ALL Operator

Unifies two input query results bez da pokusa da otkloni duplikate iz rezultata.

Znaci da   
SELECT country FROM HR.Employees  
UNION ALL   
SELECT country FROM HR.Employees

Ce vratiti ***m+n*** broj rezultata

### INTERSECT Operator

Ovo je u sustini presek skupova. Ovaj operator vraca samo redove koji se pojavljuju u ***query\_1*** i ***query\_2***

Sam INTERSECT naravno ima implicit DISTINCT.

### INTERSECT ALL Operator

U sustini postoji u SQL standardu, alil nije implementiran u T-SQL.

Alternativa:

Znaci “ALL” predstavlja da vraca sve redove bez da se otklanjaju duplikati. U slucaju INTERSECT ALL to znaci da ce se vratiti svi redovi iz preseka, ukljucujuci i duplikate.

Koriscenjem ROW\_NUMBER funkcije mozes brojati occurrences svakog reda u svakom input query-u. U sustini specificiras sve atribute u PARTITION BY clause funkcije i koristis (SELECT <any constant>) u ORDER BY clauseu funkcije.

SELECT  
ROW\_NUMBER() OVER(PARTITION BY country, region ORDER BY (SELECT 0)) AS rownum, country, region FROM HR.Employees

INTERSECT

ROW\_NUMBER() OVER(PARTITION BY country, region ORDER BY (SELECT 0)) AS rownum, country, region FROM Sales.Customers

A close-up of a map

AI-generated content may be incorrect.

Da bi otklonio **rownum** to mozes sa WITH statementom

**WITH INTERSECT\_ALL  
AS (intersect\_query\_od\_gore)  
SELECT <sve\_sem\_rownum> FROM INTERSECT\_ALL**

### EXCEPT Operator

Operator koji implementira “minus” iliti razliku skupova.

U sustini vraca redove koji se pojavljuju u PRVOM QUERY-U, ALI NE I U DRUGOM. Ovo je prvi set operator u kom redosled deklarisanja operatora postaje bitan jer jelte, samo redovi iz 1. query-a se vracaju.

### EXCEPT ALL Operator

Pazi sada ovo.

U teoriji kada imas neki red R koji se pojavljuje u query1 i red X koji je njegov corresponding row u query2. Ukoliko je broj redova R vezi od broja redova X, EXCEPT ALL ce vratiti sve one redove R koji nemaju svoj corresponding occurence u drugom query-u.

Recimo imas x10 R redova i x9 X redova. To znaci da ces dobiti x1 R red u rezultatu :)

T-SQL naravno ne nudi built in EXCEPT ALL operator, ali alternativa je takodje sa ROW\_NUMBER

A screenshot of a computer

AI-generated content may be incorrect.

### Precedence

SQL definise “precedence” izmedju skupovnih operacija.

Ides:  
INTERSECT > UNION > EXCEPT

Koriscenjem ALL varijante nista ne menja. Ukoliko se u upitu koristi vise skupovnih operacija, prvo se evaluiraju INTERSECT pa onda svi union, pa onda svi except.

SELECT \* from q1  
EXCEPT   
SELECT \* from q2  
INTERSECT  
SELECT \* from q3

Prvo se evaluira q2 i q3 iako je INTERSECT posle njih. Potom ide EXCEPT. Najveci precedence u svemu imaju zagrade pa njih mozes koristiti kako bi kontrolisao redosled operacija.

### Circumventing unsupported logical phases

Query-i koji ucestvuju u skupovnim operacijama smeju da sadrze sve klazule sem ORDER BY.

Jedina klazula koja je dozvoljena nad rezultatom skupovne operacije je ORDER BY. :)

Mozes koristiti WITH ili inline query kako bi primenjivao i druge clauses nad rezultatom skupovne operacije.

### Exercises do 230. stranice

## T-SQL For Data Analysis

Ovaj deo se fokusira na T-SQL for data analysis

## Window functions

Window function je fn koja za svaki row computeuje scalar result na osnovu kalkulacija nad podskupom redova underlying query-a.

Subset of rows(podskup redova) se naziva *window* i bazira se na window descriptoru koji se odnosi na trenutni red.

Sintaksa window functiona koristi clause ***OVER*** u kom se provideuju window specifikacije.

Primer neceg slicnog su SUM/COUNT/AVG, slucaj kada imas skup i kada trebas da vratis jedan(skalarni) rezultat za ceo skup.

U slucaju grouped query-a gubis nesto bitno – detalje. Dobijas jelte agregirane informacije, ali nakon sto grupises redove, sve komputacije u upitu moraju da se izvrsavaju nad definisanim grupama.   
Cesto moraces izvrsavati kalkulacije koje ukljucuju detail i aggregate elemente. Window function nisu tu ogranicene.   
Window function je evalueted per detailed row i primenjuje se nad subset of rows koji potice iz query result seta.   
Rezultat window funkcija je skalarna vrednost koja se dodaje u drugu kolonu u query resultu. Unlike grouped functions, windows functions don’t cause you to lose detail.

Recimo da zelis da vratis vrednost porudzbina i procenat koji svaka porudzbina zauzima od totalne sume porudzbina nekog kupca. Window funkcijom mozes vratiti customer total, zajedno sa detaljima redosleda vrednosti i mozes i computeovati procenat trenutne porudzbine naspram totalne sume svih porudzbina.

Sto se tice subqueries(self contained/correlated) – njih mozes koristiti da primenis scalar agregat calculation nad skupom, ali njihova pocetna tacka je svez pogled nad podacima, rather that the undelying query result set.

Ako underlying query ima table operators, filtere i druge elemente upita, oninece uticati sta subquery vidi. Ako zelis da subquery “gleda” na underlying query, moras ga ponoviti.

Kada su u pitanju WINDOW FUNCTIONS, window function se primenjuje nad podskupom redova underlying query result set-a. Ne nad svezim pogledom na podatke. Tako da, sve sto dodas u underlying query ce se automatski primeniti na sve window funkcije koje se koriste u query-u. Naravno, mozes restrictovati window ako zelis.

Jos jedan benefit window funkcija je sto mozes definisati order kao deo specifikacije kalkulacije sto ne prkosi relacionom pogledu na rezultate. Redosled(order) ovde se definise kao kalkulacija i ne treba se mesati sa prezentacionim smislom redosleda.

SELECT empid, ordermonth, val,  
SUM(val) OVER(PARTITION BY empid  
 ORDER BY ordermonth  
 ROWS BETWEEN UNBOUNDED PRECEDING   
 AND CURRENT ROW) AS runval  
FROM Sales.EmpOrders;

Ovo je primer querya nad Sales.EmpOrders koji ce koristiti window aggregate fn da computeuje running total values za svakog employeea i mesec.

A table of numbers with numbers

AI-generated content may be incorrect.

Postoje x3 dela u definiciji window funkcije  
OVER -> window-partition clause, window-order clause & window-frame clause

Prazan “OVER()” predstavlja ceo underlying query’s result set. Onda sta god da dodas window specifikaciji ce u sustini restrictovati window.

**Window-partition clause(PARTITION BY)** restrictuje window na subset redova koji imaju istu vrednost u particionim kolonaka kao trenutni red. U gore primeru radilo se particionisanje na osnovu *empid*. Za underling row sa empid 1, window exposeuje function filteru samo one redove gde je employee id = 1.

**Window-order clause(ORDER BY)** definise redosled(ali ne prezentacioni redosled). U window ranking funkciji, window ordering daje znacaj ranku. U nasem primeru, ordering se desava na osnovu *ordermonth* kolone.

**Window-frame clause(ROWS BETWEEN <top delimiter> AND <bottom delimiter>)** – filtrira frame ili podskup redova iz window particije izmedju dve specificirane vrednosti.   
U nasem primeru je frame definisan bez low boundary pointa (UNBOUNDED PRECEDING) i extenduje se do trenutnog reda(CURRENT ROW). In addition to the window-frame unit ROWS, postoji i RANGE, ali nije skroz implementirano u T-SQL.

NOTE: Posto pocetna tacka window funkcije je underlying query result set, a underlying query result set se evauluira tek u SELECT fazi query-a, window funkcije smeju da se koriste samo u SELECT i ORDER BY clauseu querya.

Najcesce ces ih koristiti u SELECT-u. Ako treba da je referujes u ranijem logical query processing fazi poput WHERE, moraces koristiti table expression.

### Ranking window functions

Koristis ranking window functions kada rankujes svaki row with respect to others unutar window-a. T-SQL supports x4 ranking fn’s: ***ROW\_NUMBER, RANK, DENSE\_RANK, NTILE***.

**SELECT orderid, custid, val,  
ROW\_NUMBER() (ORDER BY val) as row\_number,  
RANK() (ORDER BY val) as rank,  
DENSE\_RANK() (ORDER BY val) as dense\_rank,  
NTILE(10) OVER (ORDER BY val) as ntile  
FROM Sales.OrderValues  
ORDER BY val**

A table of numbers with text

AI-generated content may be incorrect.

ROW\_NUMBER dodeljuje inkrementalni sekvencijalni integer na svaki red u query-u na osnovu window orderinga.

Nisam bas najbolje razumeo. RANK & DENSE\_RANK je za rangiranje valjda.

RANK reflects trenutni red koji ima manji ordering value od trenutnog reda. DENSE\_RANK reflects count distinct ordering values-a.  
NTILE je kada zelis da povezes redove u rezultatu sa tileovima(equally sized groups of rows) dodeljivanjem tile numbera svakom redu. Specifikujes broj tilesova koje zelis i window ordering.

Sample query ima 830 redova i request je za 10 tiles. 830/10 => 83. Window ordering se bazira na val column. To znaci da tih 83 reda sa najmanjim vrednostima ce biti dodeljeni vrednost 1. Oni sa najvisim vrednostima tile-u 10.

Svaka od ovih window funkcija podrzava i PARTITION BY koji da si stavio, onda umesto da se primenjuje na “sve redove”, primenjivace se samo na subset redova koji je u partitionu za taj trenutni red koji se “izvrsava”.

Window funkcije se logicki izvrsavaju kao deo SELECT-a, pre DISTINCT clausea. Zasto je to bitno? Trenutno OrderValues view ima 830 redova od kojih je 795 distinct vrednosti.

**SELECT DISTINCT val, ROW\_NUMBER() OVER(ORDER BY val) as rownum  
FROM Sales.OrderValues;**

**A table of numbers and rows

AI-generated content may be incorrect.**

Rezultat je 830 redova jer se ROW\_NUMBER procesira pre DISTINCT clause-a.

Alternativa bi bila da grupises sada redove na osnovu tog polja nad kojim radis ordering jer za svaki val ce se napraviti rownum.

**SELECT val, ROW\_NUMBER() OVER(ORDER BY val) as rownum  
FROM Sales.OrderValues  
GROUP BY val;**

To ce da ti vrati 795 redova jer GROUP BY se izvrsava pre SELECT-a. Naravno, nema potrebe vise za DISTINCT.

### Offset window functions

Mozes koristiti offset window functions da bi vrati element iz reda koji je na odredjenom offsetu od trenutnog reda sa pocetka ili kraja window frame-a.

***LAG, LEAD, FIRST\_VALUE, LAST\_VALUE***

LAG & LEAD supportuju window particije i windor-order clauses. Nema relevance sa window framingom. Ove fn koristis kako bi dobio element iz reda koji je u odredjenom offset-u od trenutnog reda unutar particije na osnovu orderinga.

**LAG** – gleda na vrednost pre trenutnog reda(looks before)  
**LEAD** – gleda na vrednost posle trenutnog reda(looks ahead)

Prvi argument funkcija je obavezan i to je element koji zelis da vratis  
Drugi argument je offset(1 if not specified)  
Treci argument je default vrednost da vratis ako nema tog reda u requested offsetu(NULL default)

**SELECT custid, orderid, val,  
LAG(val) OVER( PARTITION BY custid ORDER BY orderdate, orderid) as prevval,  
LEAD(val) OVER(PARTITION BY custid ORDER BY orderdate, orderid) as nextval  
FROM Sales.OrderValues  
ORDER BY custid, orderdate, orderid**

A table of numbers with black text

AI-generated content may be incorrect.

Koristis FIRST\_VALUE & LAST\_VALUE da vracas elemente koji su prvi/poslednji iz window frame-a, pa s toga podrzavaju window partition, window order & window frame clauses.

**SELECT custid, orderid, val,  
FIRST\_VALUE(val) OVER( PARTITION BY custid ORDER BY orderdate, orderid ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) as firstval,  
LAST\_VALUE(val) OVER(PARTITION BY custid ORDER BY orderdate, orderid ROWS BETWEEN CURRENT ROW AND UNBOUNDED FOLLOWING) as lastval  
 FROM Sales.OrderValues  
 ORDER BY custid, orderdate, orderid**

A table with numbers and letters

AI-generated content may be incorrect.

SQL Server 2022 je ubacio NULL Treatment clause sa offset window functions:

**<function>(<expression>) [IGNORE NULLS | RESPECT NULLS] OVER(<specification>)**

### Aggregate window functions

Aggregate window functions se koriste za agregiranje redova unutar definisanog “window-a”. Podrzavaju window-partition, window-order & window-frame clauses.

SELECT orderid, custid, val,  
SUM(val) OVER(PARTITION BY custid) as customer\_total  
SUM(val) OVER() as grand\_total  
FROM Sales.OrderValues

U cemu je fora. Sum(val) OVER() exposeuje window nad svim redovima underlying query result set-a. Znaci svi redovi u Sales.OrderValues.   
Kada radis PARTITION BY, onda dobijas window na samo one redove koji dele istu vrednost za tu neku kolonu(custid u ovom slucaju) naspram vrednosti reda koji invokeuje tu window funkciju.

A table of numbers with a white background

AI-generated content may be incorrect.

Evo recimo query koji predstavlja npr procenat trenutno vrednosti porudzbine naspram grand total i customer total-a

SELECT orderid, custid, val,  
100. \* val / SUM(val) OVER() as pctall,  
100. \* val / SUM(val) OVER(custid) as pctcust,  
FROM Sales.OrderValues;

A screenshot of a computer

AI-generated content may be incorrect.

Ako dodamo npr window frame i window order

**SELECT empid, ordermonth, val,   
 SUM(val) OVER(PARTITION BY empid   
 ORDER BY ordermonth   
 ROWS BETWEEN UNBOUNDED PRECEDING   
 AND CURRENT ROW) AS runval   
FROM Sales.EmpOrders;**

A table of numbers and numbers

AI-generated content may be incorrect.

Kada zadajes neki window-frame, ti mozes doci i odrediti koliki frame zelis na deterministican nacin. To znaci da kazes “prethodna 2 reda i narednih 1 red”

ROWS BETWEEN 2 PRECEDING AND 1 FOLLOWING

### WINDOW Clause

Window clause ti dozvoljava da imenujes ceo window specification ili deo window spec-a u queryu i onda ga koristis u OVER clause-u window funkcija u tom queryu.

Poenta ovog clausea je da se skrati duzina querya. Dostupno je tek od sql server 2022 i higher.

Recimo da imas query koji koristi iste specifikacije nad aggregate funkcijama.

A black and white text on a white background

AI-generated content may be incorrect.

To se ponavlja x4 puta. Ono sto mozes jeste da kazes sledece  
SELECT empid, ordermonth, val  
SUM(val) OVER W as runsum,  
MIN(val) OVER W as runmin,  
MAX(val) OVER W as runmax,  
AVG(val) OVER W as runavg  
FROM Sales.EmpOrders  
WINDOW W AS (PARTITION BY empid ORDERBY ordermonth ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)  
ORDER BY custid, orderdate, orderid

Mozes i extendovati taj window

SELECT custid, orderid, val,   
FIRST\_VALUE(val) OVER(  
 PO ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS firstval,   
LAST\_VALUE(val) OVER(  
 PO ROWS BETWEEN CURRENT ROW AND UNBOUNDED FOLLOWING) AS lastval FROM Sales.OrderValues   
WINDOW PO AS (PARTITION BY custid ORDER BY orderdate, orderid)   
ORDER BY custid, orderdate, orderid;

Deli se ista particija i window-order, ali je drugaciji window-frame.

Za definisanje vise windowa

...  
WINDOW P AS (PARTITION BY custid),  
 PO AS (PARTITION BY ...),  
 POF AS (PARTITION BY ...)

## Pivoting data

Pivoting data podrazumeva rotiranje podataka iz stanja “row”-a u stanje “column”-a. Postoji i mogucnost agregiranja podataka.

Pivoting se cesto izvrsava za svrhe data reportinga od strane prezentatcionog sloja. Nekada to mozes i kroz db da odradis.

Recimo da imas sada ovakvu tabelu.

A screenshot of a computer code

AI-generated content may be incorrect.

Da moras sad da bratis total order quantity za svakog employee i customera uradis sledece

SELECT empid, custid, SUM(qty) as sumqty  
FROM dbo.Orders  
GROUP BY empid, custid  
A table with numbers and letters

AI-generated content may be incorrect.

Recimo da moras da uradis total quantity po zaposlenom(redovi) na svakog kupca(kolone)

A screenshot of a number

AI-generated content may be incorrect.

Svaki pivoting request ukljucuje x3 logicke processing faze:  
1. **Grouping** phase sa associated grouping ili na row elemente  
2. **Spreading** phase sa associated spreading ili na cols elemente

3. **Aggregating** faza sa associated aggregation elementima i aggregate funkcijama

Imagine a table of orders where each row includes an employee ID, a customer ID, and a quantity.  
If you want to see how much each employee sold to each customer:

* Group the data by **employee ID**.(empid is the grouping element)
* Turn each **customer ID** into its own column (spread).(spreading quantities by customerid)
* Use **SUM(quantity)** to fill in the values. (you need to identify the aggregate function in the aggregation element(qty attribute) which is SUM in this case)

The chapter shows how to do this using:

* A **manual grouped query**, and
* The **PIVOT table operator**, which simplifies the process.

### Pivoting with a grouped query

Ova solucija je straightforward.   
Grouping phase se radi sa GROUP BY  
Spreating phase je sa SELECT clause uz CASE expression za svaki target column. Zbog case expressiona moras da znas spreading values u napred. Jer cemo spread po empid naspram customerida(a,b,c,d), onda ce to izgledati ovako:  
CASE WHEN custid = ‘A’ THEN qty END

Taj izraz u sustini vraci quantity iz trenutnog reda samo kada je trenutni row za order customera A. Ukoliko ne specificiras ELSE, onda se vraca NULL.

Ako ne znas spreading values ahead of time, onda mozes queryovati te podatke i iz toga konstrusati query string.

Krajnji expression koji daje rezultat za customere:(dodaj jos c i d customere)

SELECT empid,  
SUM(CASE WHEN custid = ‘A’ THEN qty END) AS A,  
SUM(CASE WHEN custid = ‘B’ THEN qty END) AS B,  
FROM dbo.Orders GROUP BY empid;

### Pivoting with the pivot operator

Radi se sa PIVOT operatorom.   
Kao table operator, PIVOT operates u contextu FROM clause-a. Izvrsava se nad source table ili table expression provided to it as its left input, pivots the data, and returns a result table.

Pivot operator ukljucuje istu logical processing fazu(grouping, spreading, aggregating) samo ne treba toliko linija koda da bi se postigla ista funkcionalnost.

Forma PIVOT operatora:

SELECT ...  
FROM <input\_table>  
 PIVOT (<agg\_function>(<aggregation\_element>)   
 FOR <spreadiing\_element> IN (<list\_of\_target\_columns>)) AS <result\_tbl\_alias>  
WHERE ...;

Spreading element ce biti custid  
Aggregation element je qty  
Aggregate function je SUM  
list of target column names A, B, C, D

Ne moras eksplicitno specificirati grouping elemente pa samim tim GROUP BY otpada. U PIVOT operatoru, on ce sam skontati grouping elemente na osnovu eliminacije. Groupin elementi su svi atributi iz source tabele koji nisu bili specificirani kao spreading elementi ili aggregation elementi.

Zato moras voditi racuna da source table za PIVOT operator nema atribute druge sem grouping, spreading & agregration elemenata!

To postizes koriscenjem table expressiona koji ce ukljuciti samo atribute koji su ti neophodni.

SELECT empid, A, B, C, D  
FROM (SELECT empid, custid, qty FROM dbo.Orderes) as D  
PIVOT(SUM(qty) FOR custid IN(A,B,C,D)) as P

## Unpivoting data

Unpivoting je tehnika koja rotira podatke iz stanja kolona u stanje rowova. U sustini ono po cemu si radio spreading, sada ce biti kao da si po tome radio grouping.

Uglavnom ukljucuje queryovanje pivoted stanja podataka i kreiranje iz svakog reda, vise result rows, svaki da razlicitom source column value.

Cesta upotreba je unpivot data koje si importovao iz spreadsheet u bazi zarad lakse manipulacije.

Kreiraj sada tabelu dbo.EmpCustOrders – predstavljace pivoted set.  
A screenshot of a computer program

AI-generated content may be incorrect.

Nerelevantni preseci(nema podataka) su takodje ukljuceni u ovom skupu i obelezeni sa NULL.

Ovaj problem se moze resiti na x2 nacina   
- APPLY   
- UNPIVOT

### Unpivoting with APPLY

Ovaj nacin ukljucuje:

* 1. Producing copies
  2. Extractinv values
  3. Eliminating irrelevant rows

Prvi korak podrazumeva kreiranje kopije za svaki source row. Jedna za saku kolonu koju nameravas da unpivotujes. To znaci da ces morati kreirati kopiju za svaku kolonu A,B,C & D. Cross join ti to resava izmedju EmpCustOrders tabelom i tabelom koja ima row za svaki customer.

Ako imas tabelu customera u bazi, onda mozes koristiti tu tabelu u cross join. Ako nemas, onda mozes kreirati virutelnu koriscenjem table-value constructor based on the VALUES clause.

**SELECT \* FROM dbo.EmpCustOrders  
CROSS JOIN (VALUES(‘A’),(‘B’),(‘C’),(‘D’)) AS C(custid)**

VALUES clause definise skup x4 reda, svaki sa jednim customer ID value.   
A table of numbers with a white background

AI-generated content may be incorrect.

Drugi korak je extracting value iz jednog od izvornih customer quantity columns(A,B,C ili D) da bi se vratila jedna value colona(recimo da je nazivamo *qty*). Moraces extractovati vrednost iz column koja predstavlja trenutni custid value.

Ako je custid ‘A’, onda qty column treba da vrati vrednost iz kolne A. Ako je B, onda iz kolone B.

A screenshot of a computer program

AI-generated content may be incorrect.

Iz tog razloga ide cross apply umesto cross join. Cross apply evaluira prvo levu stranu i onda primenjuje desnu stranu za svaki levi red i na taj nacin ce se moci pristupati elementima s leve strane u desnoj strani.

CROSS JOIN => CROSS APPLY

A screenshot of a test

AI-generated content may be incorrect.

Treci korak je sto imamo sada nullove za qty. Nema poente da ih drziimo. E sad to sto je CROSS APPLY izvrsen u toku FROM, znaci da njegove kolone su nam dostupne u WHERE i mozemo samo dodati filter da se te empty kolone ne uracunaju

SELECT empid, custid, qty  
FROM dbo.EmpCustOrders  
CROSS APPLY (VALUES (‘A’, A), (‘B’,B), (‘C’,C), (‘D’,D)) AS C(custid, qty)  
WHERE qty IS NOT NULL;

A table with numbers and letters

AI-generated content may be incorrect.

### Unpivoting with the UNPIVOT operator

SELECT ...  
FROM <input\_table>  
UNPIVOT(<values\_column> FOR <names\_column> IN (<source\_columns>)) AS <result\_table\_alias\_> WHERE...

Unpivoting podrazumeva kreiranja x2 result columns iz bilo kog broja source kolona. Jedna koja drzi source column names kao stringove i druga koja drzi source column value. U ovom primeru, moras unpivot source kolone ABC & D, producing the result names column custid i qty.

SELECT empid, custid, qty  
FROM dbo.EmpCustOrders  
UNPIVOT(qty FOR custid IN (A,B,C,D)) AS U;

UNPIVOT je implementiran kao table operator u kontekstu FROM clause-a. Izvrsava se nad source tabelom(ili table expression). Unutar zagrada unpitovt operatora, specificiras ime koje zelis da assignujes koloni koja ce drzati source column values i ime koje zelis da assignujes koloni koja ce drzati source column names(custid) i listu source column names(A,B,C & D). Posle zagrada, potom deklarises alias table resultu iz tog UNPIVOT table operatora.

UNPIVOT ce automatski ukloniti NULLS, nije optional kao unutar APPLY operatora.

## Grouping sets

U sustini query sa GROUP BY clause

SELECT SUM(qty) as sumqty FROM dbo.Orders; je takodje grouped set iako nema GROUP BY

Recimo da nad istom tabelom zelis da napravis uniju podataka nad razlicitim grupacijama.

A white screen with black text

AI-generated content may be incorrect.

Prvi veliki problem ovde je performance query-a. Drugi problem je sto postoji vec native clause za ovakve stvari

### GROUPING SETS subclause

U sustini pojacava GROUP BY clause. Mozes ga koristiti da bi definisao visestruke grouping setove unutar istog querya.

SELECT empid, custid, SUM(qty) AS sumqty  
FROM dbo.Orders  
GROUP BY  
 GROUPING SETS  
 (  
 (empid, custid),  
 (empid),  
 (custid),  
 ()  
 );

Poslednji grouping set => () je u sustini “nista” tj empty grouping set. Predstavlja grand total.

Ovo je logicki ekvivalentn onom gore queryu sa puno UNION-a, ali mnogo bolje se interno optimizuje i cistije je.

### CUBE subclause

Abbreiated way to define multiple grouping sets.

U zagradi CUBE subclause-a ti postavis skup clanova razdvojene zarecom i uzmes sve moguce grouping skupove koji se mogu definisati na osnovu input membera

Npr CUBE(a, b, c) je ekvivalentno sledecem:  
**GROUPING SETS( (a,b,c), (a,b), (a, c), (b,c), (a), (b), (c), ())**

U teoriji, skup svih podskupova elemenata koji se mogu kreirati iz odredjenog skupa. Naziva se power set.

**SELECT empid, custid, SUM(qty) as sumqty  
FROM dbo.Orders  
GROUP BY CUBE(empid, custid);**

Onaj gore query bi ovakav bio sa CUBE.

### ROLLUP subclause

Takodje nudi skracen(abbreviated) way to define multiple grouping sets.

ROLLUP ne produceuje sve moguce gruouping setove poput CUBE, ROLLUP zauzme hijerarhiju unutar input clanova i kreira samo grouping sets koji formiraju leading combinations of the input members.

**ROLLUP(a,b,c)** bi kreirao x4 grouping seta gde je hijerarbhija a>b>c tj kao:   
**GROUPING SETS( (a,b,c), (a,b), (a), ())**

Na primer da zelis da vratis total quantities za sve grouping sets koji mogu biti definisani baziranjem na one time hierarchy of order year, order month, order day.

A screenshot of a computer code

AI-generated content may be incorrect.

A white paper with numbers and letters

AI-generated content may be incorrect.

### GROUPING & GROUPING\_ID functions

Kada imas single query koji definise vise grouping setova, mozda bi morao treba da uvezes rezultate redova i grouping setova.

Dokle god su svi gruping elementi definisani sa NOT NULL, to je lagano.

SELECT empid, custid, SUM(qty) as sumqty  
FROM dbo.Orders  
GROUP BY CUBE(empid, custid)

A table of numbers with black text

AI-generated content may be incorrect.

NULL o kolonama ovde je u sustini placeholder koji kaze da ta kolona nije bila ukljucena u grupisanju skupa.

E sada, ukoliko grouping column dozvoljava nullove(custid, sumqty, empid), onda nemas bas precizan nacin da razlikujes redove u kojima zaista nema vrednosti ili kolona samo nije bila ukljucena u grupisanju skupa.

GROUPING fn je jedan od nacina da se to razresi. U sustini ona prihvata ime kolone i ukoliko je element “aggregate element” onda vrati 1, a vrati 0 ukoliko je u pitanju detail element(member of the current grouping set).

U sustini malo kontraintuitivno, ali it is what it is.

A screenshot of a computer program

AI-generated content may be incorrect.

I sada ne moras vise da se oslanjas na null da bi povezao vezu izmedju result rows i grouping sets.

GROUPING\_ID funkcija moze da ga malo pojednostavi. U sustini das funkciju sa svim elementima koji su ukljuceni u grouping set kao input : GROUPING\_ID(a, b, c, d) i fn vrati integer bitmap u kom svaki bit predstavlja razliciti input element.

GROUPING\_ID function above, the grouping set (a, b, c, d) is represented by the integer 0 (0×8 + 0×4 + 0×2 + 0×1). The grouping set (a, c) is represented by the integer 5 (0×8 + 1×4 + 0×2 + 1×1), and so on

A screenshot of a computer program

AI-generated content may be incorrect.

integer 0 predstavlja grouping set (empid custid) (00)  
Integer 1 predstavlja grouping set (empid) (01)  
Integer 2 predstavlja grouping set (custid) (10)  
Integer 3 predstavlja grouping set () (11)

## Time series

Time series data su podaci koji predstavljaju “seriju” dogadjaja(series of events) illi mera(measurements) koje su uzete na odredjenom vremenskom intervalu.

Primer: Podaci o temperaturi i vlaznosti na svaki sat. Time series data analysis cesto ukljucuje organizovanje podataka po grupama(**buckets**) i onda agregiranje istih po bucketu.

**Sample data:**

Recimo imamo senzor temperature i vlaznosti koji hvata podatke intervalno. Podaci ce se cuvati u x2 tabele – Sensors(podaci o senzorima) & SensorMeasurements(ono sto senzori iscitaju)

A screenshot of a computer

AI-generated content may be incorrect.

Stranica 263. su gde se ovi podaci nalaze.

### DATE\_BUCKET Function

**DATE\_BUCKET(datepart, bucektwidth, ts[,origin])**

Azure sql edge, MSSQL 2022 & Azure SQL database

Poenta funkcije je da vrati **pocetnu tacku time bucket-a** koji sadrzi input timestamp.

Rezultat funkcije moze se koristiti kao identifier of the containing bucket i gruping element u queryu.

Kako bi identifikovao containing time bucket za neke input timestamp(ts), prvo moras razmisliti o “streli vremena” koja se deli u time buckets. Znaci neophodna ti je pocetna tacka bucketa.

Ta pocetna tacka se definise sa *origin* inputom koji moze biti bilo koji date and time data type i opcion je. Ako nije specified, onda je default 1. jan 1900.

*datepart* specificiras date and time part koji zelis da koristis poput year/month/day/hour/second i sl.

*buckektwidth* specifyujes bucket width u smislu spcificiranog dela(part). Npr ako kazes *datepart: hour & bucketwidth: 12* ti definises 12 hour buckets. Input *ts* je input timestamp za koji hoces da vratis start time za containing bucket. On moze biti bilo koji date and time i trebao bi da se matchuje sa inputom origina(po pitanju tipa). Function’s return type je isti kao i tip inputa za *ts*.

A screenshot of a computer

AI-generated content may be incorrect.

Prva x4 12h bucketa imaju sledeca start vremena:  
2022 01 01 00 05 00 -> 2022 01 01 12 05 00 -> 2022 01 02 00 05 00 -> 2022 01 02 12 05 00

Input timestamp 2022-01-02 12:00:03 je u x3 bucketu i stoga unkcija vraca vrednost:  
 2022-01-02 00:05:00 koja je start tog containing bucketa.

### Custom computation of start of containg bucket

stranica 268 ako si zainteresovan.

### Applying bucket logic to sample data

Tipicno zelis da bucketizeujes time series stored u tabeli poput *SensorMeasurements*   
Recimo zelis da saljes upite nad tebelom i za svako citanje, coputeujes respective start of the bucket.

Prvo definises 12 hour buckets starting ad midnight i onda mozes koristiti ponoc bilo kog datuma kao origin.

DECLARE  
 @bucketwidth as INT = 12  
 @origin as DATETIME2(0) = ‘19000101 00:00:00’;

SELECT sensorid, ts, DATE\_BUCKET(hour, @bucketwidth, ts, @origin) as bucketstart  
FROM dbo.SensorMeasureements

input ovoj funkciji sada je kolona “ts” iz tabele.

A screenshot of a computer

AI-generated content may be incorrect.

I sada ono sto cesto zelis da uradis jeste da grupises podatke po bucketu, a ne samo da ih podelis i onda vrsis agregate funkcije nad grupama.

Npr napravis onaj gore da je CTE uz WITH i onda  
WITH c as (...)  
SELECT sensorid, bucketstart,   
DATEADD(hour, @bucketwidth, bucketstart) as bucketend,   
MIN(temperature) as mintemp,  
MAX(temperature) as maxtemp,  
AVG(temperature) as avgtemp  
FROM c  
GROUP BY sensorid, bucketstart  
ORDER BY sensorid, bucketstartA black and white text with numbers

AI-generated content may be incorrect.Dobijes ovako nesto.

### Gap filling

Ovo gore je neki happy scenario gde manje vise vecina podataka je popunjena. Recimo da je jedan senzor bio offline i nema podataka/citanja, onda ces u sustini morati nekako popuniti te “prazne buckete”.

Npr zelis podatke o vremenu od 9. juna 2022 do 15. juna 2022 po 12h intervalima.

Agregiranjem “praznih bucketa” gde nije bilo podataka ce rezultat biti NULL.

Gap-filling je cesto hangled organizovanjem tabele koja drzi sve tacke koje cemo posmatrati(podatke). U nasem slucaju to bi znacilo sve moguce bucket start times u nedelji.

Potom bi apply left outer join izmedju te tabele i bucketized podataka sa missing bucketima. Na taj nacin dobijas sve bucket informacije ukljucujuci i missing buckets.

Referentna tabela moze biti base tabela koja se rucno kreira ili table expression.

A screenshot of a computer

AI-generated content may be incorrect.

Ukratko, bata je kreirao time series od pocetnog do krajnjeg datuma.

Mozes sad koristiti multi-CTE sa CTE za svaki korak.

Na stranici 277 imas bas objasnjenje o tome.

## Data Modification